Università degli Studi di Cagliari
International Ph.D. in Innovation Sciences and Technologies
Cycle XXXI

Innovative information models to capture the dynamics of clinical processes: introducing process-oriented traceability in medical informatics specifications as a case study

Scientific Disciplinary Sector: ING-INF/06

Ph.D. Student:	Francesca Frexia
Coordinator of the Ph.D. Programme:	Prof. Roberto Orrù
Supervisors:	Prof. Riccardo Scateni
	Gianluigi Zanetti

Final exam. Academic Year 2017 – 2018
Thesis defence: January - February 2019 Session

Abstract

Historia magistra vitae, or at least it tries; in fact, only if recorded and studied from different perspectives, the past can help people to improve their present and build a better future, avoiding mistakes that have already occurred and foreseeing the optimal directions to follow. Traceability, supporting the reconstruction and analysis of what has happened, thus becomes a way to improve safety, quality, efficiency and accuracy in all the sectors in which it is applied, particularly in medicine. The advancement of technical and scientific progress can have a positive contribution, as increasingly accurate and affordable devices are made available for the tracing of care paths. At the same time, however, to create faithful reality descriptions from the heterogeneous information collected, it is essential to study and develop methodologies relating the data to the structure, context and dynamics of the processes that generated them. This work strives to address these issues, with the definition of two theoretical information models describing process-oriented traceability in widely implemented specifications for medical informatics. The first model describes the lifecycle of a biological sample in a diagnostic process, both for clinical routine and research activities: the Specimen Event Tracking Profile is a contribution to the Pathology and Laboratory Medicine (PaLM) Domain of the Integrating the Healthcare Enterprise (IHE) guidelines. The second result is a model describing traceability for a clinical process in general, in the context of openEHR specification.

Acknowledgements

I will try to be concise and organized, "itemising" my gratitude. I want to thank:

- my employer, CRS4, and the University of Cagliari, and in particular Prof. Roberto Orrù, Prof. Giacomo Cao and Prof. Riccardo Scateni, for giving me the opportunity to work on this Ph.D.;
- my supervisor, Gianluigi, for his guidance during all these years, the stimulating discussions, the constructive disputes;
- my reviewers, François Macary and Ian McNicoll, for the trust they have shown in agreeing to review my work and for the invaluable help of their advice and corrections;
- the IHE PaLM Domain Committees, for all their support in developing the SET Profile;
- Prof. Wil van der Aalst, for the permission to use the images from his work, and for his work too;
- Luca P. e Giovanni D., for their useful tips as predecessors;
- Alessandro and Cecilia, my companions in the wonderful journey in the worlds of IHE and openEHR;
- Anna and Riccardo, for encouraging me to embark on this adventure;
- Paolo, for all the inspiring and consolatory discussions and because, with his splendid family, he adopted Nerino;

- my friends at work (Vittorio, Luca L., Massimo, Marco Enrico, Francesco, Mauro, Emanuela, Katia, Andrea), because they are always ready to help me and working with them is engaging and (almost always) fun;
- my friends around the world (Luisa, Elena, Francesca, Christine, Adriana, Anna, Stefano, Cosimo, Daniela, Monica, Paola, Andrea, Marco, Matteo, Stefania, Antonello, Martina, Luca), because I'm happy I have met them;
- my wanderlast friends (Valeria and Roberta), who have proven to know me very well, with the right advice at the right time;
- my "camp" friends (Luana, the Irenes, Giorgia, Orazio and Paola), for their encouragement and respect for the work to be done, in every free second (or almost);
- the god of the reckless, because his unreliable help sometimes pulls me out of complicated situations, but his inconstancy pushes me to count on myself first of all;
- Pippo, Zorro and Micetta, because only my cats really understand me;
- my parents, that have taught me most of what it is really important to know;
- Ivano and Stefania, the mainstays of this family;
- Chiara, Davide and Giacomo (in order of appearance), because they play with me and treat me like a rock star.

Dedication

To those who care.

Contents

Chapter 1

Introduction

1.1 Research Context and Motivation

The advancement of science and technology is profoundly transforming the majority of the human activities. Very common operations, such as paying for a purchase, are now usually intermediated by systems and devices. This trend is also significant in the clinical practice, since most of the treatment paths involve technological support in one or more phases of the process, such as: the management of the patient during hospitalization with Admission, Discharge, Transfer systems; the carrying out of an examination by means of complex medical devices (CT scan, Magnetic Resonance Imaging, etc.); the distribution of drugs by means of unit dose systems; the long term evaluation of a patient clinical conditions, on the basis of previous records in hospital information systems; the remote control of the state of health via monitoring devices. Medical research too is deeply correlated with technological development, since the creation of more precise and cheaper tools paves the way to attempt to answer questions that could not be dealt with before. The continuous improvement of data acquisition technologies also has a positive impact in the traceability of a medical - health or research – process. The presence of reliable and inexpensive sensors enables, in fact, the acquisition of context data, in addition to those directly related to the ongoing activ-

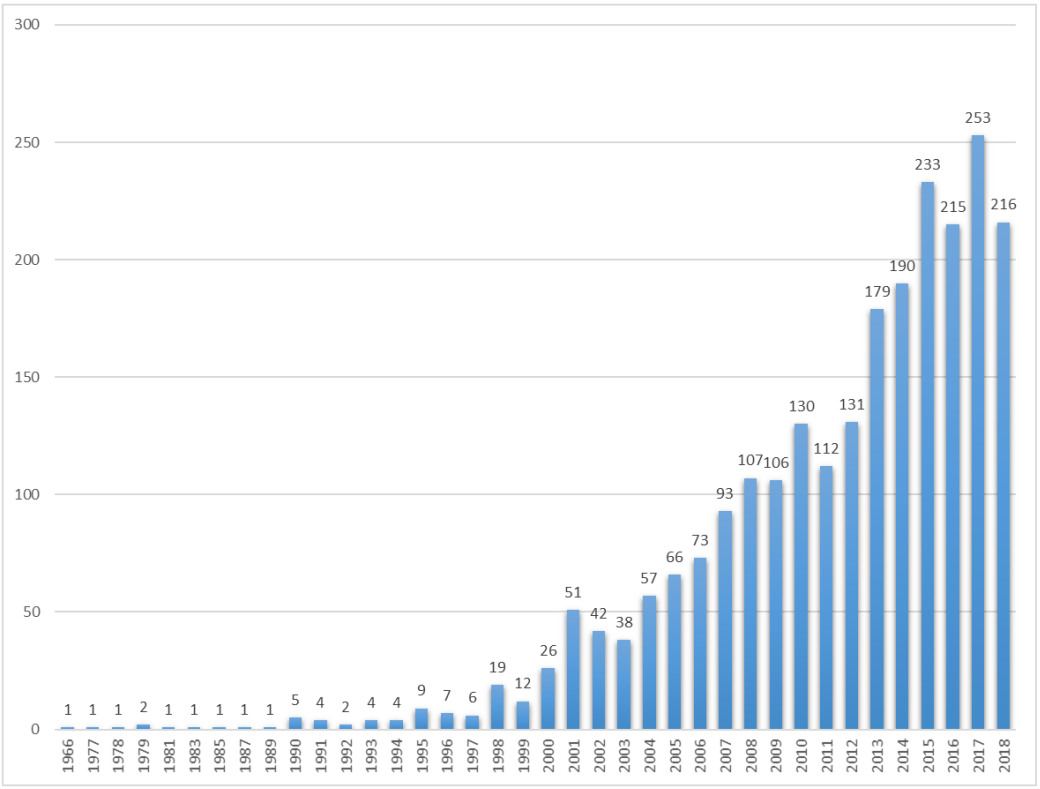

Figure 1.1: **Traceability in Medical Publications.** Number of publications about traceability in the PubMed database from 1966 to 2018. *[data set downloaded in September 2018].*

ity. Traceability has in particular gained increasing attention in the medical sector, where it has been adopted for example in supply chain improvement, error prevention or reduction, public health, legal medicine, clinical research, epidemiological surveillance, process management and billing [8].

The interest in traceability in healthcare shown by institutions, industry and research is growing, as can be sensed by the acceleration in the production of scientific publications centred on traceability shown in Fig. 1.1. As traceability applications multiplying, the lack of standards in the field is resulting in a multitude of custom data formats and models, giving rise to widespread interoperability problems among the methodologies used to collect and analyse the data. As a result, several healthcare organizations and research groups are trying to implement initiatives for traceability data modelling. The aim of this work is the definition of new information models to describe traceability in clinical processes and medical research, preserving semantics and

supporting interoperability.

1.2 Research Scope

The following points define the scope of the research activities described in this thesis.

- The research is focused on introducing clinical process traceability in the specifications issued by Integrating the Healthcare Enterprise (IHE) [56] and openEHR [33], two of the most important international bodies for the definition of guidelines in the field of healthcare IT. The decision is motivated by the fact that these are the best formalisms supporting the use of consolidated standard in process-oriented guidelines developed by domain experts - IHE - and ensuring a robust long-term preservation of clinical semantics - openEHR.
- The specific context is traceability in medicine, for healthcare practice and biomedical research, with a particular focus on the workflows for pathology and laboratory medicine. This choice is related both to the dramatic impact of laboratory results on patient diagnostic/therapeutic pathway and to the special attention in laboratory medicine to traceability issues, denoted, for instance, by the existence of metrological traceability for laboratory analytical methods.
- The results are theoretical models. No reference implementation has been realized as the models are developed as part of standard formalisms and will undertake a process of presentation, evaluation, revision and approval before they can be used in practice.

1.3 Research Contributions

This work addresses the issues previously described with two original contributions, briefly outlined below.

- **IHE Specimen Event Tracking Profile**. An IHE compliant information model capturing traceability information during a specimen lifecycle in a pathology or clinical laboratory examination. The idea

underpinning the profile was approved by the Technical Committee of the IHE Pathology and Laboratory Medicine Domain and the result presented in this thesis depicts the status of the profile development, realized with the contribution of the Committee itself, during the periodic meetings envisaged by the development cycle of the IHE guidelines.

- **openEHR Traceability Model**. A formalism for an openEHR-consistent representation of process-oriented traceability in clinical pathways, complementing the openEHR Task Planning Model recently introduced in openEHR. This idea and the model described in this thesis have not yet been submitted to the openEHR community.

1.4 Thesis Structure

The remainder of this thesis is organized as follows:

- **Chapter 2** provides some background information about traceability in medicine;
- **Chapter 3** presents a state of the art on process modelling, with particular attention to the IHE and openEHR specifications, which are the main context of the work of this dissertation;
- **Chapter 4** describes the contribution of this thesis to the IHE specifications, the *IHE Set Event Tracking Profile*;
- **Chapter 5** presents the proposed model for introducing process-oriented traceability into the openEHR specifications, *the openEHR Traceability model*;
- **Chapter 6** summarises the contributions and sets out the expected future developments.

Chapter 2

Background: Traceability in Medicine

2.1 Introduction

The definition of "traceable" in Oxford Dictionary [88] is *able to be followed on its course or to its origin*: it is very generic, but it clearly conveys the broad spectrum of meanings and interpretations that can be brought together under the aegis of "traceability". There is, in fact, on the one hand, an absence of unanimity on the definition of traceability [76] and, on the other hand, a difficulty in finding definitions that are not recursive, such as "traceability is the ability to trace..." [60] [61] [59] [81] or "traceability is the ability to track" [32] [6] [51]. The term, first recorded in 1740–50 [45], appeared in the agribusiness sector in the midst of the 1990s [13], but a strong public and institutional interest has certainly been fuelled by a series of food scandals, as the Bovine Spongiform Encephalopathy disease [101] or the Hudson Foods recall of 25 million pounds of ground beef at the behest of the U.S. Department of Agriculture [52]. The food industry was therefore the first sector in which tracking tools, methods and standards were disseminated, shortly followed by other areas, as the potential of tracking was understood and explored. At present, traceability is widespread in applications related, for examples, to:

- supply chain, to control the logistics aspects of distribution;

Figure 2.1: **Perspectives in traceability.**

- manufacturing, to govern production and, recently, also to show the level of ecological sustainability in process of realization of the goods;
- accounting, to track pieces of financial information and assign costs to activities;
- quality control, to trace and evaluate the adequacy of activities or products.

This Chapter is focused on the declinations that traceability can assume in medicine, identifying on three main dimensions, as shown in Fig.2.1:

- **traceability of things**, analogous to internet of things, related to the tracking of physical objects during their creation, distribution, use, etc.;

- **traceability of information**, related to the control of the data created in process evolution or product development, analysing their provenance to evaluate their accuracy and quality;
- **traceability of processes**, concerning the overall workflow, to identify and follow activities, actors, timings in the general framework of the complete process.

This distinction is only functional to the treatment of a wide and multifaceted subject: the three dimensions are not to be considered mutually exclusive, on the contrary they are integrated in every traceability application.

2.2 Traceability of Things

In medicine, even more than in other fields, the "traceability of things" is closely linked to safety: medicaments, prosthesis, implantable devices, human products, diagnostic equipments are directly connected to patients and a problem in the production or in the distribution process could lead to severe consequences for health conditions. Tracking, in addition to being a tool for improving process logistics, is mainly seen as a way to enable a quick recall of dangerous or defective products from the market and alert patients of relevant problems. Technology can support the traceability of drugs and devices [84]: particular attention is dedicated to high risk and high cost medications [83].The extreme criticality has led government authorities and international bodies to monitor the entire supply chain, from manufacture to use by citizens. For Medical Products of Human Origin (MPHO), the World Health Organization has solicited member States "to encourage the implementation of globally consistent coding systems to facilitate national and international traceability" [79]. In the United States, the control function is operated by the Food and Drug Administration (FDA), the agency within the U.S. Department of Health and Human Services "protecting the public health by assuring the safety, effectiveness, quality, and security of human and veterinary drugs, vaccines and other biological products, and medical devices" [25]. The FDA orders manufacturers to implement a tracking system for devices *"whose failure would be reasonably likely to have serious, adverse health consequences; or which is intended to be implanted in the human body for more than one year; or are life-sustaining or life-supporting devices used outside of a device user facility"* [26]. The required tracking information elements are identified, but manufacturers can use different tracking methods,

documenting them in written standard operating procedures (SOP). A list of devices to be tracked is defined [27], but FDA can discretionally add tracking orders for devices outside the list or release devices from tracking. Device traceability at the moment is mainly based on the device batch number [96] and incomplete barcode data [17], but it will be improved by the use of a Unique Device Identification (UDI) System, adopted both by FDA [28] and by European Commission [18]. UDI is *"intended to provide a single, globally harmonized system for positive identification of medical devices"* [31]. The system has three main components, which will appear on the label template, as shown in Fig.2.2:

- a unique code, numerical or alphanumerical, to be conveyed by using Automatic Identification and Data Capture technology (AIDC, like barcodes, RFID, smart card, etc.) and formed by:
  - a mandatory Device Identifier (DI), which identifies the specific version or model of a device;
  - a conditional Product Identifier (PI), which depends on the manufacturer and can have data about production lot or batch, device serial number, device creation or expiration date;
- a database (UDID), which in Europe is the European Databank on Medical Devices (EUDAMED), in the US is Global Unique Device Identification Database (GUDID);
- a code system providing unique codes for every device models, to be used in data exchanges, the Global Medical Device Nomenclature (GMDN).

The Global Traceability Standard for Healthcare (GTSH) [92] is a process standard, providing a framework which describes the traceability in supply chains. GTSH defines a minimum set of shared requirements for all stakeholders, independent from their countries, technologies, organisation size or operational sophistication. It is maintained by the Healthcare Division of GS1, an international, not-for-profit association that creates and implements standards for supply chains across industries.
In the near future, for traceability of objects it is planned to explore the potential of distributed ledger and blockchain technologies, to improve the supply chain, with a special focus on the pharmaceutical supply, medical devices and Internet of Healthy Things (IoHT). The underpinning idea is that, thanks to their capability of high integrity tracking, these technologies could

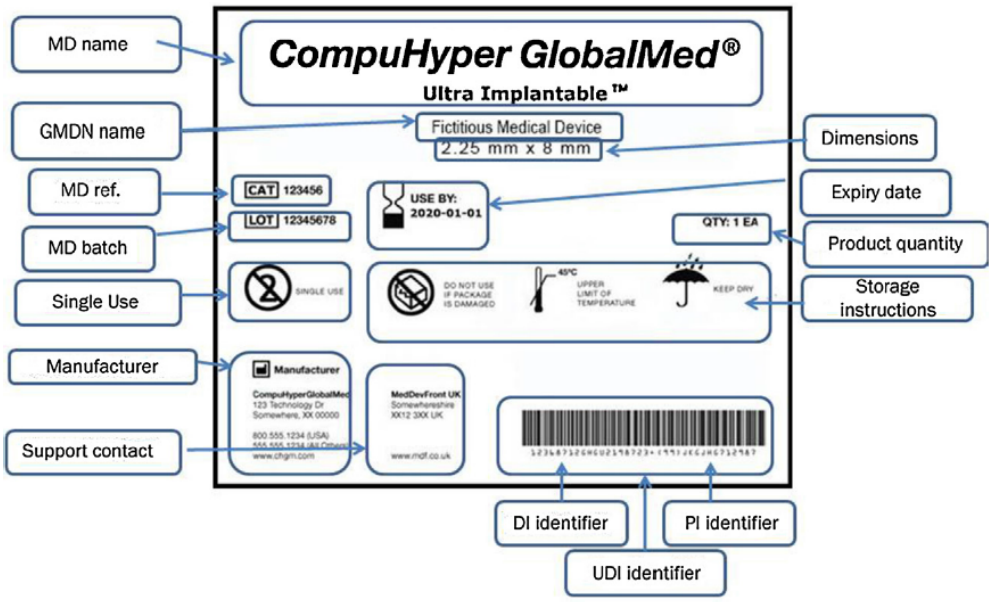

Figure 2.2: **Example of UDI label for devices, according to new specifications.** [96]

respond to the growing demand for traceability and transparency [20] [46], but still a number of issues need to be addressed, both technical and organisational [29].

2.3 Traceability of Information

Clinical data availability is a very important element both in a research study and in the choice of the best care plan for a patient [72], as a limited access to health information is considered a relevant cause of errors and inefficiency [19] [67]. At the same time, it is equally important to have a clear assessment of data accuracy: "traceability of information" is intended as mainly associated to the concepts of quality and reliability and, in this dimension, metrological traceability and provenance analysis are two major aspects. The first is, in general, a *"property of a measurement result whereby the result can be related to a reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty"* [41]. Laboratory medicine is one of the clinical fields where metrological traceability is most present and has also become a regulatory requirement, with the imple-

mentation of the European Union Directive on In Vitro Diagnostics (IVD) devices [80]. The basic idea is that measurement results must be linked to commonly accepted references, so that they can be comparable across different measurement systems, locations and times [99]. Metrological traceability consists of three main steps: establishing a reference system (reference methods + reference materials); calibrating measurements procedures according to the previously defined reference system; verifying the consistency of measurements, by comparing sets of real patient specimens to control the uniformity of results using different methods. There are two principal approaches, standardization and harmonization: standardization assesses traceability to the SI, harmonization to a reference system conventionally agreed [98]. Several national and international organisations and institutions issue guidelines and reference procedures, such as the International Organization for Standardization [40] [1], the Clinical Laboratory Standards Institute (CLSI) [14], the Joint Committee for Traceability in Laboratory Medicine [95] [63], the International Consortium for Harmonization of Clinical Laboratory Results [49], the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) [15], the American Association of Clinical Chemistry (AACC) [2]. Also IVD manufacturers are involved in best-practices refinements, being required to apply the theory of metrological to assay calibration schemes [5].
Another important aspect of information traceability is provenance, which allows data recipient to trace information back to its source, supporting the assessment of the trustworthiness of different sources [68]. For health data, tracking provenance is particularly difficult, due to the continual reworking and re-processing of clinical information, throughout the diagnostic process and into reporting analysis [24]. Provenance consists in the capture and analysis of metadata associated to data creation and processing, related to the state of digital entities at different phases of their lifecycle. Two pillar concepts are granularity and layering, which refer to the size of the basic entities and to the layer of the software stack associated to provenance metadata [10]. Provenance is also one of the FAIR Principles to improve the Findability, Accessibility, Interoperability, and Reuse of digital assets. The basic idea is that *"for others to reuse your data, they should know where the data came from (i.e., clear story of origin/history, see R1), who to cite and/or how you wish to be acknowledged"* [50]. Well established provenance models in information technology are the Open Provenance Model [78] [74]and W3C PROV [100] [75], which provides the high level structure of provenance records shown in Fig. 2.3.

In healthcare can be cited ISO/TS 18308 for provenance requirements in openEHR design principles [57], HL7 FHIR Provenance [53] for clinical

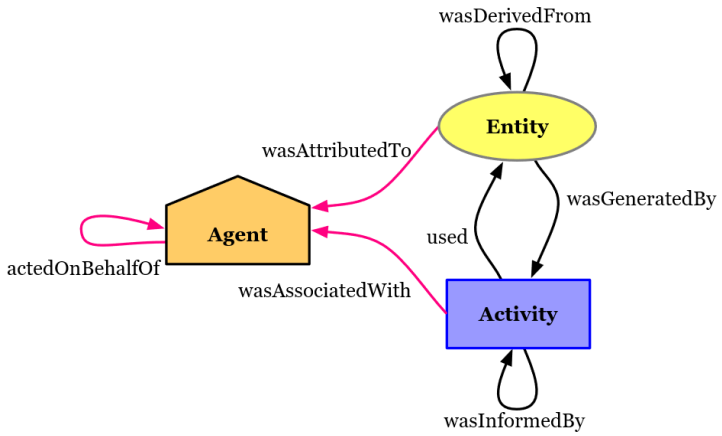

Figure 2.3: **High level overview of the structure of PROV records.** [100]

data, CEN/TC 140 [12] for molecular IVD examinations and SPREC [8] for pre-analytical quality parameters. An ISO pre-PWI proposal in biomedical research is under development for provenance information management [62]. The interest for provenance in the medical field is on the rise, fuelled by the growing awareness that as healthcare routine data are used in clinical research, integrating provenance data from these systems would lead to more complete and reliable results [48].
Another aspect of the traceability of information worth mentioning, even if it is not specific to medicine, is that relating to software engineering. In particular, for what is about requirement traceability defined as "the ability to describe and follow the life of a requirement in both a forwards and backwards direction". [37] Requirement traceability is specially relevant when developing safety-critical systems, as in the case of some medical devices in healthcare [89].

2.4 Traceability of Processes

The process dimension of traceability analyses what happened, to understand why it happened in order to predict what will happen and to move towards the best that can happen: it is therefore strictly related to efficiency and accountability, in the sense that it reconstructs cause-effect relations linked to the phases of the process, in order to improve the process itself. This approach considers traceability not only as the recording/detection of a series

of events, but as a tool to improve process modelling and management. In information technology, ISO defines a *Traceability reference model for business transactions* [58], based on five principles:

1. unambiguous identification;
2. record-keeping;
3. defined level of granularity;
4. ensuring traceability among parties;
5. precision in temporal and/or location referencing.

In the clinical domain, process traceability is mainly related to the end-to-end evaluation of clinical pathways, in hospitals and healthcare networks. It is again laboratory medicine that pays particular attention to traceability, this time in the process-related dimension. In past decades, accountability has been improved, for example, thanks to a systematic evaluation of the causes of errors directly or indirectly associated with the clinical laboratory process. This anlysis has led, for example, to verify that a large part of the error causes resided in specific problems not so much of the analytical phase itself, but of the pre- and post-analysis phases [86] [85] [11]. For the efficiency perspective, a series of reports [44] [42] by the Task Force on the Impact of Laboratory Medicine on Clinical Management and Outcomes (TF-ICO) [16] inspired several studies showing that the measure of the impact of laboratory services for patient outcome is not reliable without considering the overall diagnostic process [91] [43]. Process-based traceability in clinical practice is also powered by the increasing diffusion of sensors and devices, which enable the acquisition of data in the different steps of the treatment paths, without requiring further commitment from the operators involved [90] [71]. These tracked events can feed processing systems – based for example on process mining or machine learning algorithms - that can remotely control the process, highlight most frequent paths, report problems during execution, suggest corrective actions or show differences in behaviour with respect to the normal configuration. As an example, Fig. 2.4 shows for a phlebotomy process the comparison between a theoretical model and the real behaviour, measured by traceability events: the numbers indicate the frequencies for each arc and each node, while the unpredicted paths are highlighted in red.

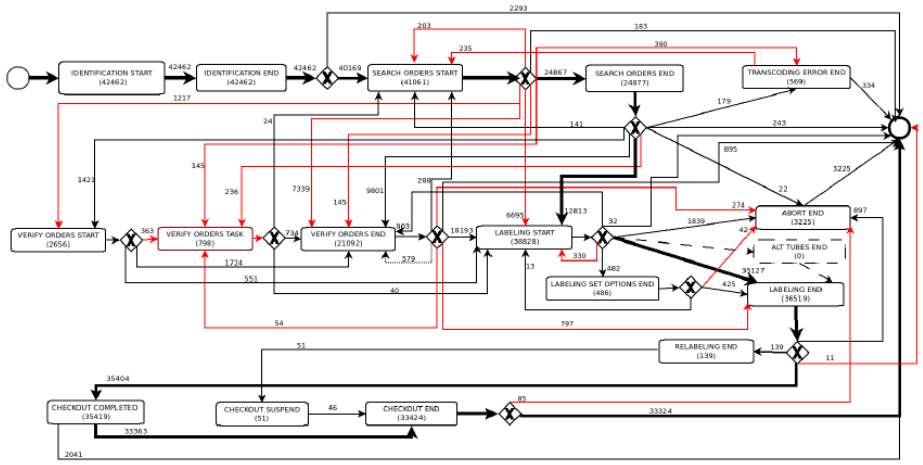

Figure 2.4: **Real process detection starting from traceability events.** [93]

2.5 Summary

Although intuitively it is a concept of immediate understanding, giving a systematic definition of traceability presents some difficulties and often requires clarifications regarding the context. In this chapter, a brief overview of some applications of traceability in medicine has been presented, introducing the distinction of three main dimensions, referring to things, information and processes. These dimensions do not identify separate categories, but provide a key reading consistent with the results presented below in this thesis, which are informative models for process-oriented traceability, in general and with a particular focus on biological samples in clinical pathways and biomedical research.

Chapter 3

Methods: the Process Perspective

3.1 Introduction

The last thirty years have seen an increasing emphasis on the process-oriented analysis of reality. Up to the 90's, in fact, information systems were almost completely data-oriented: information technology was aimed at collecting and storing data, and systems were designed on the basis of data models. The need for additional tools to capture the dynamics of real processes has subsequently led to the development of "process aware" information systems, capable of correlating data to the specific organizational context [4]. Workflow Management (WFM) and Business Process Management (BPM) were developed to fulfill this need to combine the data-centric perspective with a vision that took into account elements such as business activities, resources, cases, tasks. The Workflow Management Coalition defines:

- Workflow as "the automation of a business process, in whole or part, during which documents, information or tasks are passed from one participant to another for action, according to a set of procedural rules" [66];
- Workflow Management Systems (WFMS) as the "system that defines, creates and manages the execution of workflows through the use of

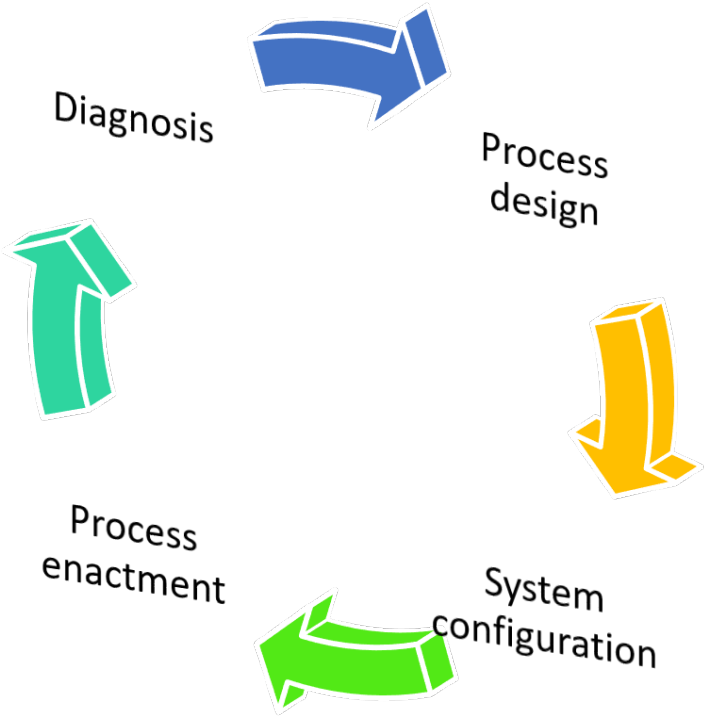

Figure 3.1: **BPM Lifecycle.**

> software, running on one or more workflow engines, which is able to interpret the process definition, interact with workflow participants and, where required, invoke the use of IT tools and applications" [66].

A WFMS combines several perspectives, such as control flow (or process) perspective, resource (or organization) perspective, data (or information) perspective, task (or function) perspective, operation (or application) perspective. Business Process Management (BPM) can be viewed as an extension of classical Workflow Management (WFM), as they face analogue business needs with different approaches: WFM supports the coordination of repetitive tasks, focusing on people role and work instructions, while BMP aims to process improvement, with a stronger emphasis on coordination than on automation. Considering the BPM lifecycle (design-configuration-enactment-diagnosis) depicted in Fig.3.1, WFM is mainly related to the lower half, while BPM also includes the upper half. This chapter presents an overview of the main process modelling notations, also introducing process mining techniques, and outlines how processes are modelled in IHE and openEHR.

3.2 Process modelling notations

Modelling a process is a complex task as "making a good model is an art rather than a science" [3], but several modelling formalisms and tools exists to support the modelling activities: the following paragraphs briefly describe some of the most used at the moment.

3.2.1 Unified Modeling Language (UML)

The Unified Modeling Language (UML) is a family of graphical notations, generally used in describing and designing software systems [35], suitable also for (business) process modelling [23]. The UML is a standard, controlled by the Object Management Group (OMG) [39], an open membership, not-for-profit computer industry standards consortium that produces and maintains computer industry specifications for interoperable, portable, and reusable enterprise applications in distributed, heterogeneous environments. UML defines diagrams which can complement each other to form a coherent view of a process. In particular interaction diagrams describe how groups of objects collaborate: sequence diagrams, for example, depicts the objects and the messages that are passed between these objects within a use case.

3.2.2 Business Process Model and Notation (BPMN)

Also Business Process Model and Notation (BPMN) is a specification developed and maintained by the OMG [38]. BPMN is one of the most widely used notation for business process modelling and is supported by many vendors. In BPMN a Process is depicted as a graph of Flow Elements, which are a set of Activities, Events, Gateways, and Sequence Flows that define finite execution semantics. Processes can be defined at any level from enterprise-wide Processes to Processes performed by a single person. Low-level Processes can be grouped together to achieve a common business goal.

3.2.3 Yet Another Workflow Language (YAWL)

The acronym YAWL stands for "Yet Another Workflow Language", which is both a workflow modelling language and an open-source workflow system [47]. The development of the YAWL language was heavily influenced

by the Workflow Patterns Initiative [102], started in 1999 with the goal of defining a conceptual basis for process technology, describing relevant patterns for one of these perspectivse: control-flow, data, resource, exception handling and event log imperfections. Based on a systematic analysis of the constructs used by existing process modeling notations and workflow languages, a large collection of patterns was identified. The aim of YAWL is to cover many patterns while keeping the language simple. YAWL modelling recalls Petri nets, bipartite graphs modelling processes in terms of places and transitions, with token flowing through the network according to firing rules. In YAWL activities are called tasks and conditions correspond to places.

3.2.4 Case Management Modelling Notation (CMMN)

Applications of Case management include licensing and permitting in government, application and claim processing in insurance, patient care and medical diagnosis in healthcare, mortgage processing in banking, problem resolution in call centers, sales and operations planning, invoice discrepancy handling, maintenance and repair of machines and equipment, and engineering of made-to-order products. CMMN is a graphical notation used for capturing work methods based on the handling of cases requiring various activities, that may be performed in an unpredictable order in response to evolving situations, like healthcare processes. CMMN defines a common meta-model and notation for modeling and graphically expressing a Case, as well as an interchange format for exchanging Case models among different tools, using an event-centered approach and the concept of a case file. CMMN is another OMG specification, intended to be consistent and complementary to BPMN [38].

3.2.5 Common Workflow Language (CWL)

The Common Workflow Language (CWL) [65] is a specification for describing analysis workflows and tools in a way that makes them portable and scalable across a variety of software and hardware environments, from workstations to cluster, cloud, and high performance computing (HPC) environments. CWL is designed to meet the needs of data-intensive science, such as Bioinformatics, Medical Imaging, Astronomy, Physics, and Chemistry.

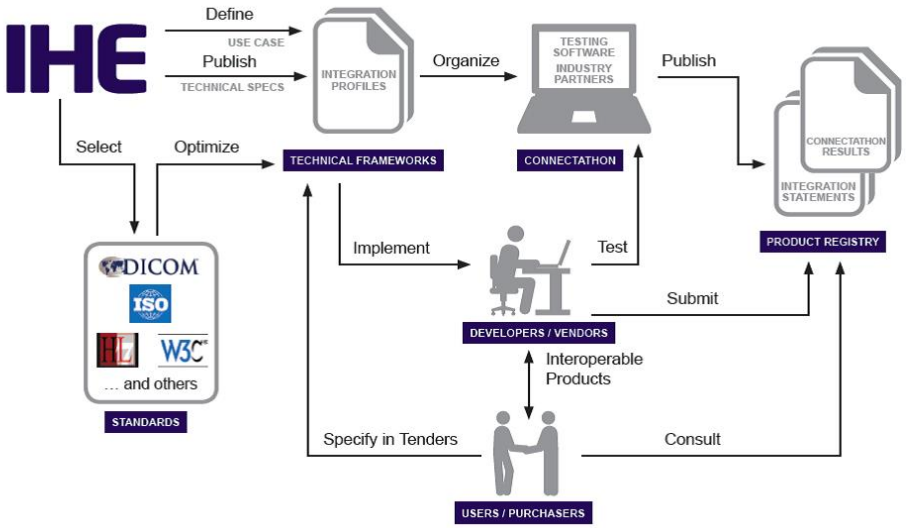

Figure 3.2: **IHE Process.** [55]

3.3 Process modelling in IHE and openEHR

3.3.1 Integrating the Healthcare Enterprise (IHE)

The Integrating the Healthcare Enterprise (IHE) [56] is the reference institution for the interoperability of information systems in the healthcare environment; the consortium is divided into clinical Domains (Cardiology, Radiology, Pharmacy, etc.), each governed by Committees, which develop and maintain the interoperability guidelines, under the form of Technical Frameworks (TF) [82]. The specifications define the required implementations of established health IT standards to achieve effective systems integration, facilitate appropriate sharing of medical information and support optimal patient care. When implemented in industrial products,TF are periodically tested at annually events called Connectathons and can be continuously improved, according to the process depicted in Fig. 3.2.

For each of the defined Domains,Technical Frameworks describe the main processes in the form of use cases, workflows, actors and transactions, mapping significant events and real practice situations. They are usually divided into three volumes:

- Volume 1, containing the high level user view of the Profiles, designed

to solve well-defined use cases, describing through a set of models, how information systems should interact to support the use cases and the associated processes. Each Profile involves a small number functional roles played by information systems, called "Actors", and defines Transactions that flow between Actors.

- Volume 2, providing the technical specifications of implementation of the standards selected to carry the Transactions between Actors.
- Volume 3, defining the content modules used in the Domain.

The Pathology and Laboratory Medicine Domain, born in 2016 from the merger of Laboratory and Anatomic Pathology Domains, is focused on the modelling of the diagnostic process related to in-vitro specimens, in terms of structured data, documents, images, scheduling, performing and reporting observation, all in a digital format. The Fig. 3.3 depicts an overview of all the PaLM Domain profiles listed belows, with an indication about their maturity:

- Laboratory Analytical Workflow (LAW) – final version, included in TF;
- Laboratory Testing Workflow (LTW) – final version, included in TF;
- Laboratory Device Automation (LDA) – final version, included in TF;
- Laboratory Point Of Care Testing (LPOCT) – final version, included in TF;
- Laboratory Code Set Distribution (LCSD) – final version, included in TF;
- Laboratory Specimen Barcode Labeling (LBL) – final version, included in TF;
- Sharing Laboratory Reports (XD-LAB) – final version, included in TF;
- Laboratory Clinical Communication (LCC) – public comment, published as a supplement to the TF;
- Specimen Event Tracking (SET) – under development, published as a supplement to the TF;
- Anatomic Pathology Structured Report release 2.0 – trial implementation, published as a supplement to the TF;

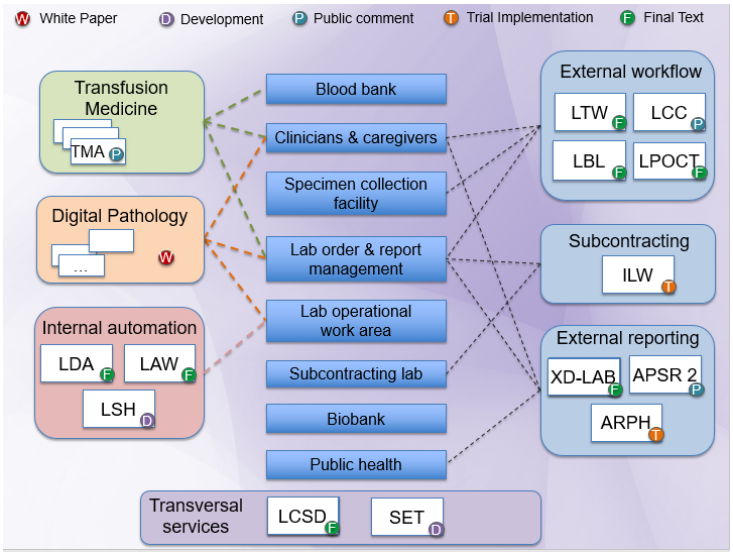

Figure 3.3: **IHE PaLM Profiles.** [82]

- Digital Pathology and Structured Reporting (SR) – white paper, published as a supplement to the TF;
- Transfusion Medicine (TMA) – public comment, published as a supplement to the TF;
- Anatomic Pathology Reporting to Public Health (ARPH) – trial implementation, published as a supplement to the TF;
- Inter Laboratory Workflow (ILW) – trial implementation, published as a supplement to the TF;
- Laboratory Specimen Handoff– under development, published as a supplement to the TF.

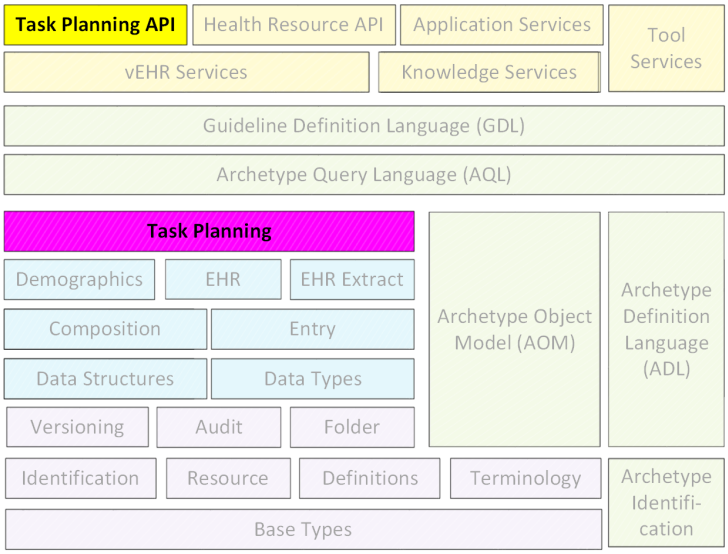

Figure 3.4: **openEHR Task Planning Model.** [30]

3.3.2 openEHR

The openEHR Task Planning model [30] has been introduced in 2016 and models clinical processes as a "journey navigation system", actualizing the guidelines created by the medical community in a workplan. This "workable conceptual model" supports the clinician in care delivery, adapting the path to the "driver" decisions, the patient conditions and external factors which can deviate the care plan from ideal guideline prescriptions.

The approach followed in creating the specification is strongly operator-centric: a process, indeed, is not modelled as a – more or less - complex machine, where the clinician is a gear with a specific role to play and a series of actions to execute. openEHR perspective on workflows is based on the consideration that health processes are strongly knowledge intensive, driven by human participants reacting to changing context, with constant exceptions and shifting goals. openEHR Task Planning model, in summary, is "an open architecture for a process-enabled EHR that helps the clinical team to take a person needing care from where they are now to a goal state, via an efficient, evidence-based pathway tailored for the patient". The specification describes the models and semantics for task plans and their execution by a notional openEHR Task Planning execution engine ('TP engine'). This is assumed to operate in a server computing environment in which other systems exist with which the TP engine communicates. A key system is the

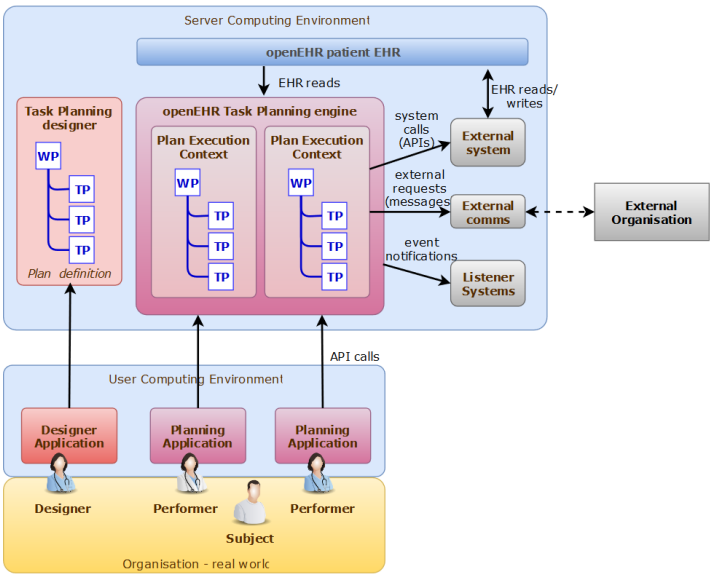

Figure 3.5: **openEHR Task Planning Engine.** [30]

openEHR EHR system, via which openEHR patient EHRs are accessible to the TP Engine. Users who act as performers of tasks are connected to the engine via applications that communicate to the TP engine through the API of the TP Engine. These applications may be dedicated only to Task Planning or TP can be a component of systems or devices in the User Computing Environment. Users who act as plan designers are connected to a task planning designer system via a dedicated designer application, as 3.5.

The paradigm identified explicitly divides planning from execution, providing:

- a detailed formalism for describing the future of a clinical process, in terms of work plans, task plan and task actions - via the definition package;
- a minimal model to deal with the present at execution time – in the materialised package, still not normative and to be completed;
- a model of audit events in the EXECUTION_HISTORY – via the history package.

The materialised structure mirrors the structure of the original definition, with the number of repetitions required by the plan activator and without

unreachable branches in the workflow graph for the execution. A materialised work plan, therefore, is not an instance of the WORK_PLAN class created at run-time but is a process representation which consist of:

- a copy of the definition form of the Plan;
- an instance of the Plan using Materialised model instances, the M_XXX classes representing the concrete executable Tasks, each with a reference back to a Task definition;
- repeatable sections 'unfolded' into as many instances as required by the Plan activator;
- removed branches for the cases in which trigger conditions or events will never be satisfied;
- Tasks preliminary allocations;
- an instance of EXECUTION_HISTORY , root point to accumulate Plan execution event records.

The materlialised Work Plan will be activated when it will be ready to be used, starting the clock for the workplan. After activation, modifications to the plan are allowed only ahead of the runtime window, which is a moving window corresponding to the section of a materialised Plan actively executing at any moment. A Task can correspond to any of openEHR ENTRY as shown on the Clinical Investigator process diagram in Fig. 3.6.

3.4 Process Mining

According to the IEEE Task Force on Process Mining Manifesto [97], the aim of process mining is "to discover, monitor and improve real processes (i.e., not assumed processes) by extracting knowledge from event logs readily available in today's (information) systems" [3]. The discipline emerged in the last twenty year and the task force was established in 2009, motivated by the growing interest in log-based process analysis. Process Mining basis include process model-driven approaches and data mining concepts, but end-to-end process centric-analytics can offer added values, creating process maps based on current data and oriented according to the stakeholders requirements: process mining can bridge the gap between Business Intelligence

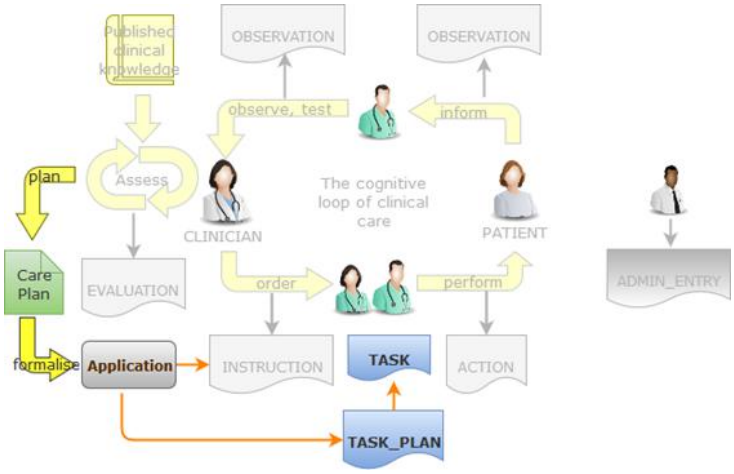

Figure 3.6: **openEHR Clinical Investigator process diagram with Task Planning.** [30]

and Business Process Management. Possible use cases are related to the discovery of process models from data, the detection of highways or bottlenecks, the creation and optimization of key performance indicators. Event logs, sequentially recorded during the process or pre-processed to obtain structured data, are the input of process mining: (re)ordered events are then correlated to activities, belonging to a process instance (case). When possible, more information can be associated to logs, as time-stamp, event initiator or other kind of data elements. The possible relationships between process models and event logs are called play-in, play-out and replay: in the first case the model is deduced from the log, in the second the real events are the input used for a simulation on the model while replay has both process model and logs as input, to check the model behavior with real data. There are three main different types of Process Mining [3]:

- discovery: Discovery consists in taking as input an event log, and then applying one or more algorithms to discover a model from that log, without using any apriori information. Event attributes, if provided, can help to obtain some different models, according to the point of view of the analysis.

- conformance: This type of Process Mining compares an existing process model with an event log of the same process. Conformance checking is able to tell us if the reality conforms to the model and vice versa, showing and quantifying all deviations.

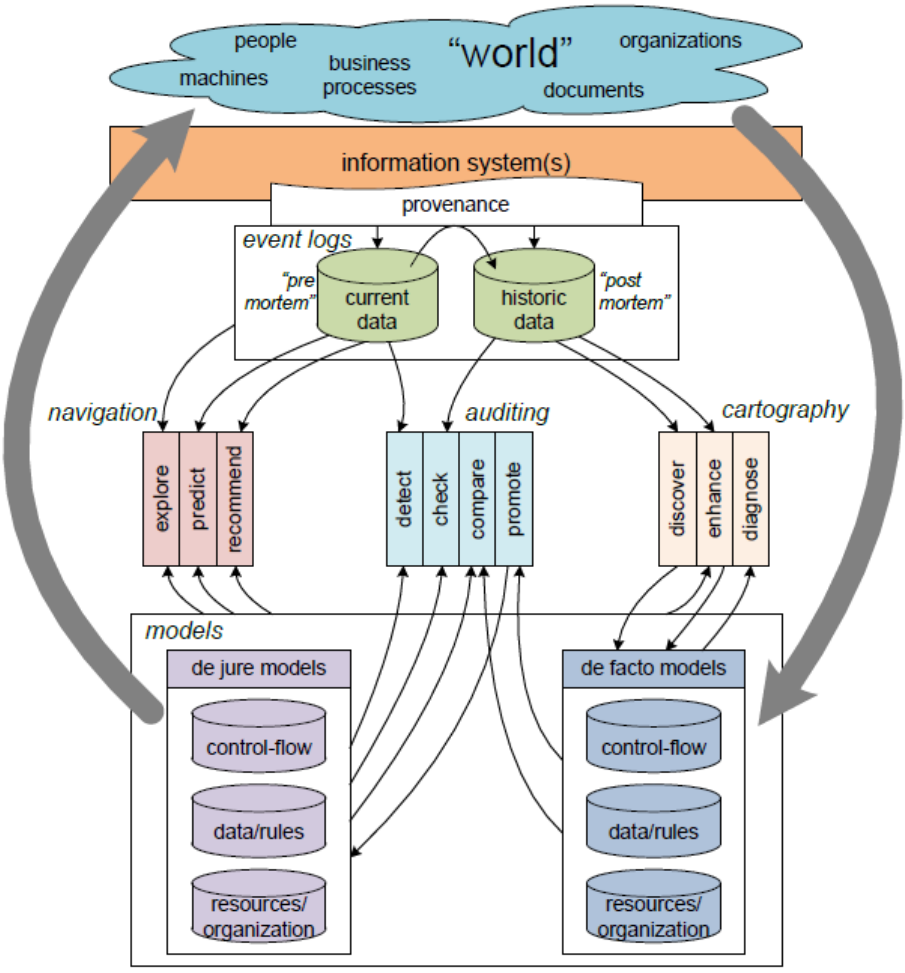

Figure 3.7: **Process Mining Framework** [3]

- enhancement: It consists in improving an existing process model using information coming from logs, showing for example bottlenecks, or suggesting some extensions to the process itself.

Fig. 3.7 shows the complete process mining framework, including ten activities grouped into three categories: cartography, auditing, and navigation.

Process Mining technique applications to the healthcare domain increased in the last decade [70], at first to discover clinical workflows from real data [69] [73], later to also evaluate the conformance to medical guidelines of actual behavior [64] [22].

3.5 Summary

Process science "combines knowledge from information technology and knowledge from management sciences to improve and run operational processes" [3], but the process perspective is not in competition with the central data approach, on the contrary the two visions are synergic for realizing a better representation of reality. This Chapter showed a brief overview on the most used process modelling notations, with a special attention to IHE and openEHR process models, which are at the starting point of the work of this thesis, whose results are described in details in the following Chapters.

Chapter 4

Introducing process-oriented traceability in IHE

4.1 Introduction

Specimens play a central role in diagnostic services, clinical care and biomedical research applications. In the clinical field, a specimen can be defined as a physical object (or a collection of objects) considered by a laboratory a "single discrete, uniquely identified unit that is the subject of one or more steps in the laboratory (diagnostic) workflow" [21], and this definition can be considered complete also from a research perspective. The data about the specimen are as important as the specimen itself, documenting both the material and the process that generated it. As such, they have an essential value for data analysis. The result of a clinical test or a biomedical research analysis is, indeed, the consequence of a series of steps, directly involved in the biological examination or supporting it: for example, the path of a laboratory examination, from the request to the response and consequent medical intervention, is defined as "a series of interrelated and interacting activities that transform a biological sample of the patient into analytical results and diagnostic information". The so-called Total Testing Process (TTP), taken from industrial literature and introduced in Laboratory in the 1970s [36], finds its best representation in the [87] [9] – constantly evolving - *brain-to-brain loop* in Fig.4.1, which encloses the practical description and the theoretical view of the generation of the laboratory medical information.

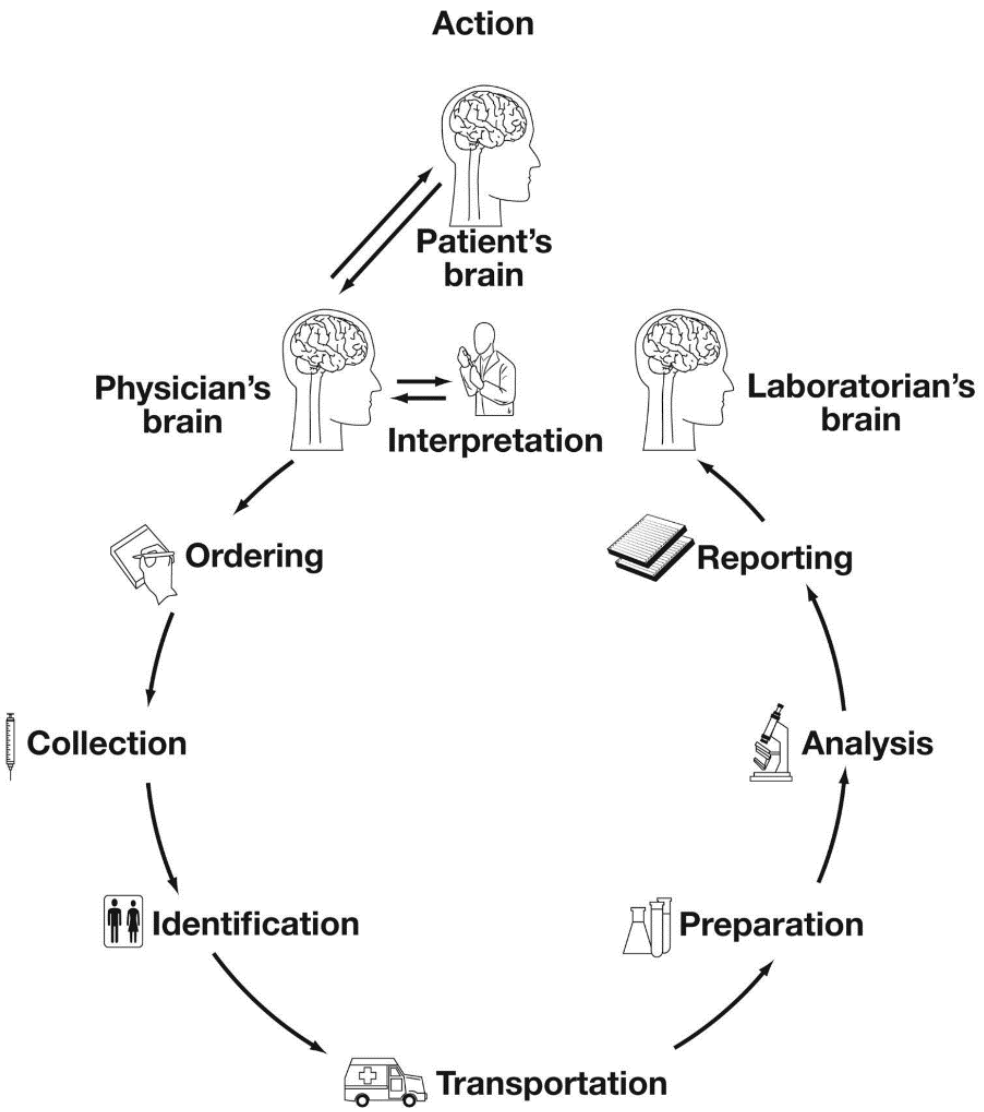

Figure 4.1: **Brain To Brain Loop.**
[87]

IHE adopts the process-oriented approach to model clinical domains, with the aim of promoting clinical systems integration and standards adoption: IHE Technical Frameworks defines use cases, workflows and transactions that can be mapped to significant events. They form a solid basis for a clinical traceability system as they are defined by a wide number of experts in the field and they can provide useful information about correctness and completeness of the whole process chain. The IHE Profiles for Pathology and Laboratory Medicine, summarized in Chapter 3, can be used to map all the clinical steps of the Total Testing Process. Each profile, indeed, models described how the actors interact during the stage of the diagnostic process, which are the trigger events and the data associated to each transaction. The proposed Specimen Event (SET) Profile covers the entire diagnostic process from a from a a new point of view, the specimen perspective, providing an information model that allows to trace the entire path of the biological sample, from collection to storage for subsequent analysis, including the transport step. This Chapter is focused on the SET Profile actors and transaction, describing how the proposed profile evolved and was developed in collaboration with the IHE PaLM Technical Committee. In particular, this Chapter contains the main elements defining the content for Volume 1 of the SET Profile, while Volume 2 is still under development.

4.2 The Specimen Event Tracking Profile

The Specimen Event Tracking (SET) Profile reads the diagnostic process in Pathology and Laboratory Medicine in a specimen-oriented perspective, defining standardised exchangeable data structures to convey the events that happened to the specimen during the analysis workflow. The Profile is focused on tracking a series of samples macro-activities such as collection, transport, storage, retrieval, high-level processing. More specific operations are not considered in scope for SET, but are treated in other PaLM Domain profiles. Each step of the process is therefore described according to the events and the relevant metadata deriving from the operations performed on the biological samples to be examined, considering how the transaction changes in a set of selected use cases. The profile belongs to the Pathology and Laboratory domain and in this scope the “specimen” concept can be referred to each kind of biological sample which can be collected for testing purpose (blood, tissue, urine, etc.). In the same way, the term “laboratory” includes generically a site where a series of examinations and tests can be performed, independently from the specialty (microbiology, hematology,

pathology, biochemistry, molecular biology). The basic idea is that the to-be-tracked steps – manual or automatic - of the process can be associated to software agents emitting events, when triggered by the operators or the devices performing the processing or analysis step for the sample. The tracking can be related to subjects in the same laboratory, in the same hospital or between different clinical institutions: in some case there will be a specific transfer phase – a shipping – performed by a courier. The workflow and the use cases identified are mainly focused on clinical routine, but part of the profile is also related to biobank samples management.

4.3 Main concepts

4.3.1 General environment

The SET Profile Actor Diagram, depicted in Fig.4.2, illustrates that the profile consists of two Actors (Specimen Event Tracker and Specimen Event Informer) and a unique Transaction (Track Specimen Information). The transaction will be called LAB-Y1, as all PaLM transactions are indicated with the code "LAB-", followed by the transaction number established when the transaction is included in the Technical Frameworks: SET Profile is still an ongoing project, so from now on the number will be substituted with the code "Y1".

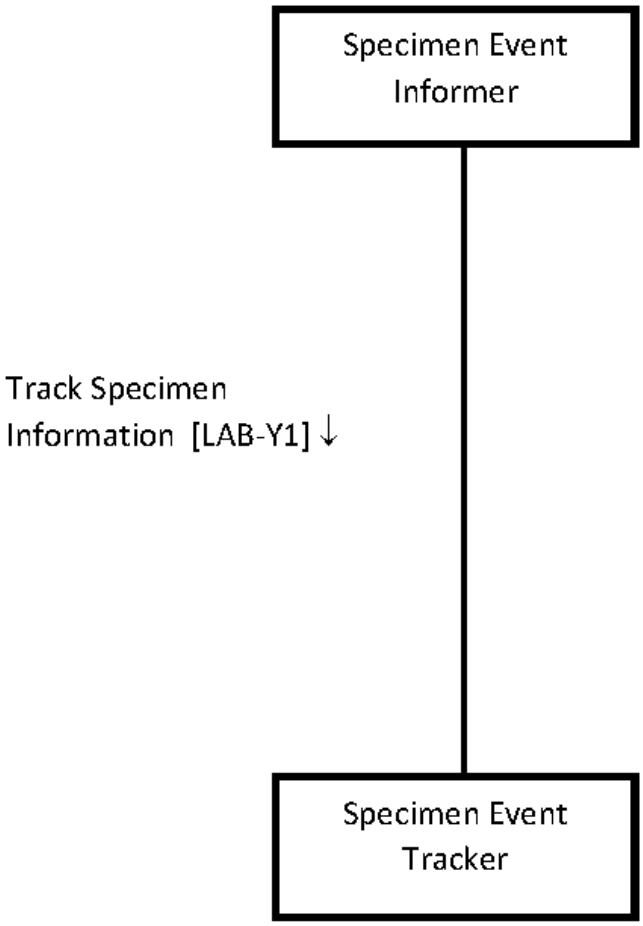

Figure 4.2: **SET Profile Actor Diagram.**

To be implemented, the Profile requires therefore the presence of (at least) a system acting like an event notification sender and (at least) a system acting like a receiver of event notification, communicating in the structured way specified by the transaction. A diagnostic process is organized, in general, as a series of operations that can be performed on a sample in a manual or an automatic way. Each macro step could have its software module acting as a Specimen Event Informer or can share it with other systems supporting the workflow. In this second case, the software module acting as a Specimen Event Informer will receive all the information about the events and will create the messages to be sent to the Specimen Event Tracker, according to the specification of the transaction. To claim compliance with this Profile, playing the role of SET or SEI, a system has to support the Track Event Information transaction, which is mandatory.

4.3.2 Actors

Specimen Event Informer (SEI)

The Specimen Event Informer (SEI) Actor triggers a message of LAB-Y1 transaction whenever a pre-defined event is recorded by the system implementing this Actor. The tracking information associated with the event is conveyed by the LAB-Y1 message. This information is coded in the definition of the events further described in details.

Specimen Event Tracker (SET)

The Specimen Event Tracker (SET) Actor collects information tracking specimen events, from messages of Transaction LAB-Y1 received from one or more SEI. The way the SET Actor processes the received information is out of scope for the Profile. For example, the SET actor could be a module of a wider traceability system, receiving the events and using the associated information to perform a run-time control of the process and to send alert messages to performers/devices involved in the diagnostic workow. But also, the SET actor could simply store the event in a database for further usage, like performance analysis or chain of custody reconstruction in case of legal problems about the execution of the clinical test.

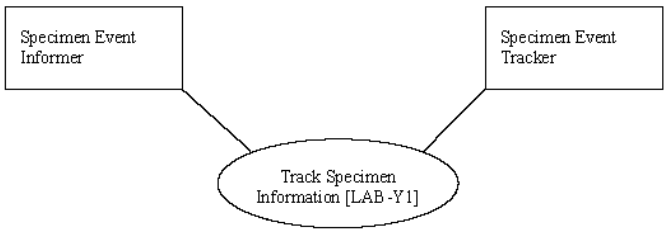

Figure 4.3: **SET Use Case Diagram.**

Actor Roles

Fig.4.3 depicts the Use Case Diagram for the SET Profile:

- Specimen Event Informer provides and sends Specimen Event Tracking events;
- Specimen Event Tracker receives and collects Specimen Event Tracking events.

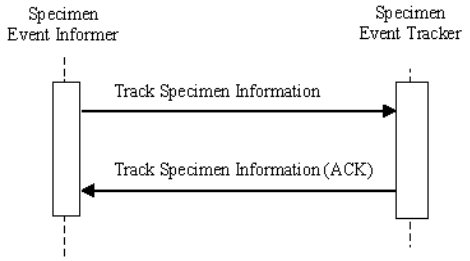

Figure 4.4: **SET Interaction Diagram.**

4.3.3 Transactions

Track Specimen Information [LAB-Y1]

For the SET Profile a single transaction - Track Specimen Information transaction LAB-Y1 - is defined to deal with the relevant information associated to the sample lifecycle. Each macro-activity is associated to an event, created or detected by the Specimen Event Informer actor, which then sends the relative messages to one or more Specimen Event Tracker actors, according to the general traceability system configuration. At message reception, the SET actor sends an acknowledgment message, which can be positive or negative (malformed event message), if the tracking event is not compliant to the related macro-activity definition. In this second case, the SEI can correct and resend. A resend can also be necessary if the SET actor system is not reachable, in case of timeout. How the tracking messages are used by the system implementing the SET actor is out of SET Profile's scope.

Table 4.1: Use case mapping to specimen life cycle steps

Specimen life cycle step	**SET use case**
Collection	Specimen containers production and collection tracking
Transfer	Specimen inter and intra organization transfer
Analysis	Intra-laboratory IVD testing specimen tracking
Derivation	Specimen derived tracking
Biobanking	Specimen biobank transfer

Trigger Events

LAB-Y1 is triggered by one of the actions associated to the tracking event list. When the action is performed, the Specimen Event Informer collects or builds the relative tracking messages. For examples, if the specimen is identified (action=identification), SEI actor will produce and send the message for identification to the SET actor. The selection of event metadata used as input reference material the HL7 Domain Analysis Model (DAM) for Specimen [54] The tracked events have a common structure, composed of a "header" part with general information and a part containing data about the specific event.

Use case definition

For the SET Profile, five use cases has been identified, described in detail in next paragraph. The guiding criteria for the selection was the attempt to cover the main steps in specimen clinical lifecycle, also adding an opening to research environments: Table 4.1 lists all the defined use cases, associating them to these stages.

Events

The SET profile aim is the tracking of a sample lifecycle, via the recording of a series of events associated with the steps of the analytical process, also including non-clinical activities such as transport or re-identification. A series of events have been selected to model the process and for each events also a set of metadata, mandatory or optional has been coded and described: every LAB-Y1 transaction message refers to a specific performed operation (storage, processing, shipping, etc.), described by one of the events in Fig.4.5.

Event	Description
Specimen Collected	The specimen has been collected by an operator (phlebotomist, physician, nurse, etc.) in a ward or in a collection room. Specimen collection action includes specimen identification.
Specimen Containers Prepared	The containers where specimens will be collected have been prepared: the appropriate containers have been selected and labeled.
Specimen Shipped	The specimen has been delivered from a location to another, which could be a different location in the same institution, or even a remote location. The specimen container could be part of a box, as many containers are transported all together to the receiving location.
Specimen Received	This event identifies the check-in of the specimen, when it arrives from an external location, and registers the successful arrival.
Specimen Accepted	This event reports the acceptance of the specimen after its check-in, meaning the specimen has passed all quality checks needed to be processed inside the receiving institution.
Specimen Rejected	The specimen has been rejected after check-in, as it does not have the minimum quality requirements to be processed ahead. This event SHALL carry the reject reason.
Specimen Re-identified	This event is tracked when the specimen needs to be re-identified by the receiving facility or institution.
Specimen Archived	The specimen has been archived in a storage system or biobank.
Specimen Retrieved	The specimen has been retrieved from a storage system or biobank, for further usage.
Specimen Derived	The specimen has been derived from a previous one. The event is useful, for example when some blocks are derived from a tissue specimen, or some slides are derived from a block. The event is also useful for biobanks, where a parent specimen is processed to create new ones for further testing.
Query for Specimens Retrieval	A query has been executed in a storage system/biobank to retrieve one or more specimens, for further usage (i.e., testing).
Specimen Sent for Testing	The specimen has been shipped to automation, or to a testing location, if no automation is present in the laboratory.
Specimen Processing Start	Specimen processing started on the referred Device
Specimen Processing End	Specimen processing ended on the referred device
Specimen Arrived at Laboratory Device	The specimen has been taken in charge by a Laboratory Device
Specimen left Laboratory Device	The specimen left a Laboratory Device
Specimen Discarded	The specimen has been discarded after all IVD tests have been performed

Figure 4.5: **SET Event List Table.**

Event	Description
Event Type	Type of the Event ("Event" column of table X.1-2)
Event Timestamp	Timestamp of event creation
Event Performer	Operator/Machine/device responsible for the event
Event ID	Unique identifier of the event
Specimen ID	Unique Identifier of the Specimen the event is related to
Specimen Container ID	Unique Identifier of the Specimen Container the event is related to
Sending Organization	Identifier of the organization sending the tracking message
Sending Facility	Identifier of the facility sending the tracking message
Receiving Organization	Identifier of the organization receiving the tracking message
Receiving Facility	Identifier of the facility receiving the tracking message

Figure 4.6: **SET Common Metadata Table.**

4.3.4 Metadata for the Events

For each event, a collection of metadata are defined, listed in the following tables: some of them are common to all the events, others are specific to the stage of the diagnostic process to which the event refers. The metadata will determine the information carried by the LAB-Y1 transaction messages, specified in Volume 2.

Event	Attribute	Data type	Description	Card.	Usage	DAM Mapping
Specimen Collected	Placer Order Number	String	PON as usual in HL7 v2	0…*	RE	No mapping, use HL7 v2 instead?
	Collector	String /ID	Person responsible of Specimen Collection	1…1	R	Performer.identifier
	Type	String/Code	Specimen Type as usual in HL7 V2	0…1	O	Specimen.typeCode
	Form	String/Code	Material nature (i.e., liquid, gas)	0…1	O	Specimen.formCode
	Description	String	Additional specimen information	0…1	O	Specimen.description
	Procedure	String	Activity performed for collection (i.e., venipuncture, biopsy)	0…1	O	SpecimenCollectionProcedure.ProcedureCode
	Coll. Date Range	Timestamp (range)	Time range of collection duration	0…1	C	SpecimenCollectionProcedure.actualCollectionDateRange
	Missed Reason?	String	Reason of collection not completed	1…1	C	SpecimenCollectionProcedure.missedReason
	Container Name	String	Name (model?) of the container	0…1	RE	SpecimenContainer.name
	Container Material	String/Code	Material of the specimen container	0…1	O	SpecimenContainer.containerMaterialCode
	Container Capacity	Number/Code	Capacity of the specimen Container	0…1	O	SpecimenContainer.Parameters.capacity
	Container Additive	String/Code	Additive of the specimen Container	0…1	O	SpecimenContainer.additive
	Container Rank	Number	Rank of the container collecting the specimen (1=first, 2= second, and so on)	0…1	RE	
	Container Height	Number	Height of the specimen container	0…1	O	SpecimenContainerParameters.height
	Expiration Time	Timestamp	Date after the specimen is no longer viable	1…1	R	Specimen.expirationTime

Figure 4.7: **SET Common Metadata Table.**

The common event metadata are general parameters related to basic information, like event timestamp or performer: these data are all required for each event. For the specific metadata, the level of detail increases and some of the fields are optional. Some of them are also associated to the relevant event in HL7 DAM Model for Specimen, to support the mapping that will be necessary to define the message structure in Volume 2 detailed description of the LAB-Y1 transaction. All the metadata selected are listed in the Appendix. Fig.4.7 presents an example for the Specimen Collection Event.

The list will be further reviewed during the development of the Transactions, as the detailed consideration of the content of the related message/s will probably lead to modifications. Examples of already known points to be fixed are:

- for the attribute "Missed Reason" or "Is De-identified" will be decided if a new event has to be created;
- the attribute "Accept Timestamp" will be present only if the event will be mapped with a shipment HL7 v2 message;
- the attributes "Involved Speciality" and "Involved Diagnosis" could be included from order information (OBR-44 segment in HL7 v2 message)

4.4 Relationship with other IHE Profiles

Each Profile in IHE PaLM Technical Frameworks cover an aspect of in vitro diagnostic process and deals, directly or indirectly, with specimen, which is encapsulated:

- for Laboratory Testing Workflow Profile, in Orders;
- for Laboratory Device Automation Profile, in Specimen Work Orders;
- for Laboratory Analytical Workflow Profile, in the Analytical Work Orders;
- for Laboratory Specimen Barcode Labeling Profile in Labeling info and labeled containers preparation;
- for XD-LAB Profile in Reports.

The SET Profile is placed in a longitudinal position with respect to the existing profiles as it explicits specimen perspective during the analysis steps, providing a global sample-oriented view of the diagnostic process. Considering the PaLM Domain ongoing projects, an apparent overlap could emerge between the operational areas of the SET profile and the Laboratory Specimen Handhof (LSH) Profile. LSH's goal is to provide standardized workflows for specimen passing between Laboratory Automation Systems and Specimen Processing Devices, starting from common experiences in the IVD industry.

Anyway, analysing in details their definition, the differences in SET and LSH Profile scopes is evident, as the first aims to trace what happened to a specimen, while the second is dedicated to provide a shared protocol for specimen management in one of the diagnostic process steps.

4.5 Use cases

4.5.1 Use Case 1: Specimen containers production and collection tracking

This use case is focused on the collection of the specimen in a collection location (ward, laboratory collecting room, surgery room). The scenario starts with the labeled containers production; this action may be performed by a robotic system or manually by an operator (identified by the generic actor Specimen Container Producer). Then, the person responsible for collecting specimens (identified with the generic actor Specimen Collector) performs the collection. It has to be noticed that when containers production and specimen collection happen at the same time, also the correspondent events may be merged into one: both container and collection information will be in this case carried by the "Specimen Collected event". The optional "Specimen Containers Prepared" event is used, instead, to cover all those scenarios where containers labeling and collection happen at different times. The SEI actor sends a LAB-Y1 message to notify the SET actor that the specimen containers have been effectively prepared, and in a second time, another message to notify the collection of the specimens.

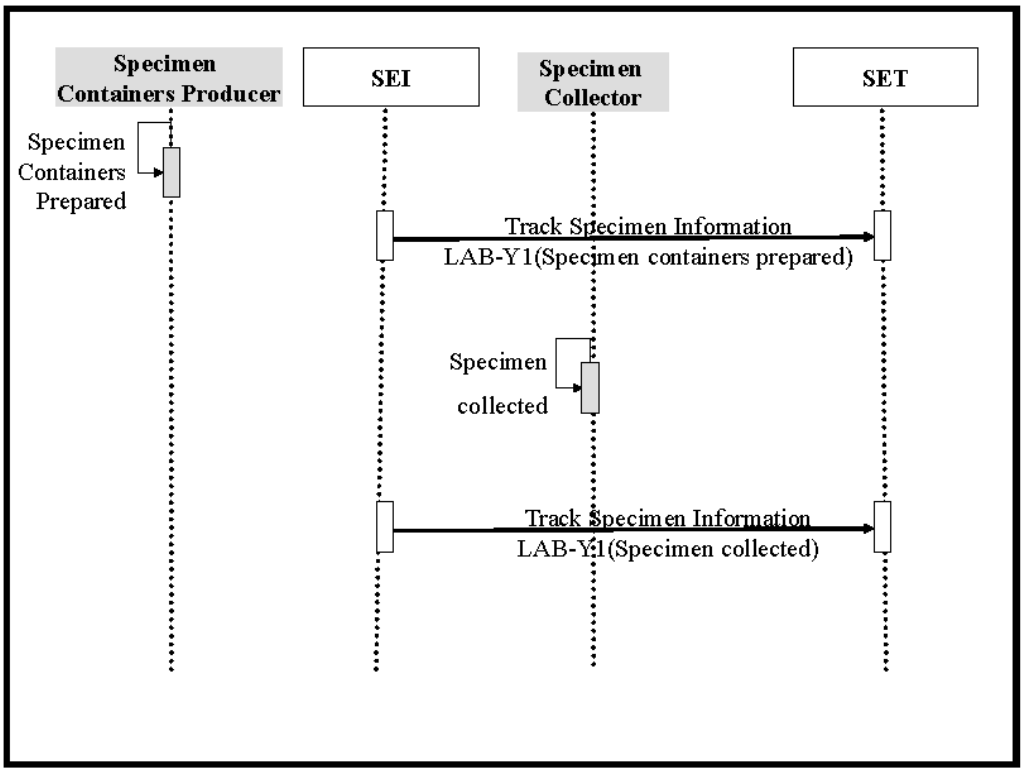

Figure 4.8: **Specimen containers production and collection tracking process flow.**

4.5.2 Use Case 2: Specimen inter and intra organization transfer

This use case involves the transfer of specimens from a shipping location to a receiving one. These two locations may belong to the same institution (intra-organization transfer) or to different institutions (inter-organization transfer). In case of inter-organization transfer, the two locations could be far from each other; such a transfer is needed when there is a subcontract between two organizations for the execution of a specific set of tests. The requester (in this use case context, the shipping location) ships the specimens, which are received and accepted by the subcontractor (in this use case context, the receiving location). Concerning the intra-organization transfer, instead, the two locations are close the one another: the specimen transfer occurs, for example, between a ward and the Laboratory of the same hospital, or different facilities belonging to the same organization. In the second case, the main difference is that transport time is usually shorter. The basic scenario starts with the specimen collection, happening at shipping location; specimens are then received, accepted and identified by the receiving location, which is also responsible for processing them for testing. It could happen that some specimen containers need to be labeled at arrival (as they travelled without any conventional label), or need to be re-identified (and so re-labeled). In some situations one or more specimens could also be refused by the receiving location (due to insufficient sample quantity, inappropriate or broken container, etc.). For these reasons, for this use case three different process flows are identified:

1. Specimen transferred, no re-identification by receiver
2. Specimen transferred and re-identified by receiver
3. Specimen rejected by receiver

The following sections analyze each sub-use case, showing the related sequence diagram. As the collecting and receiving locations could refer to different organizations, the diagrams suppose that each organization has its own SEI and SET actors. This is only one of the possible architectures, as the same diagrams may be depicted with a unique instance for SEI and SET, as it would be common to happen in case of intra-organization transfer. The diagrams show each single event being sent to both SET actors. There might be variations to that: some of the events might interest only one of the

two organizations (for example, a re-identification event might interest only the receiver's local SET). The diagrams only refer to an example of possible architecture.

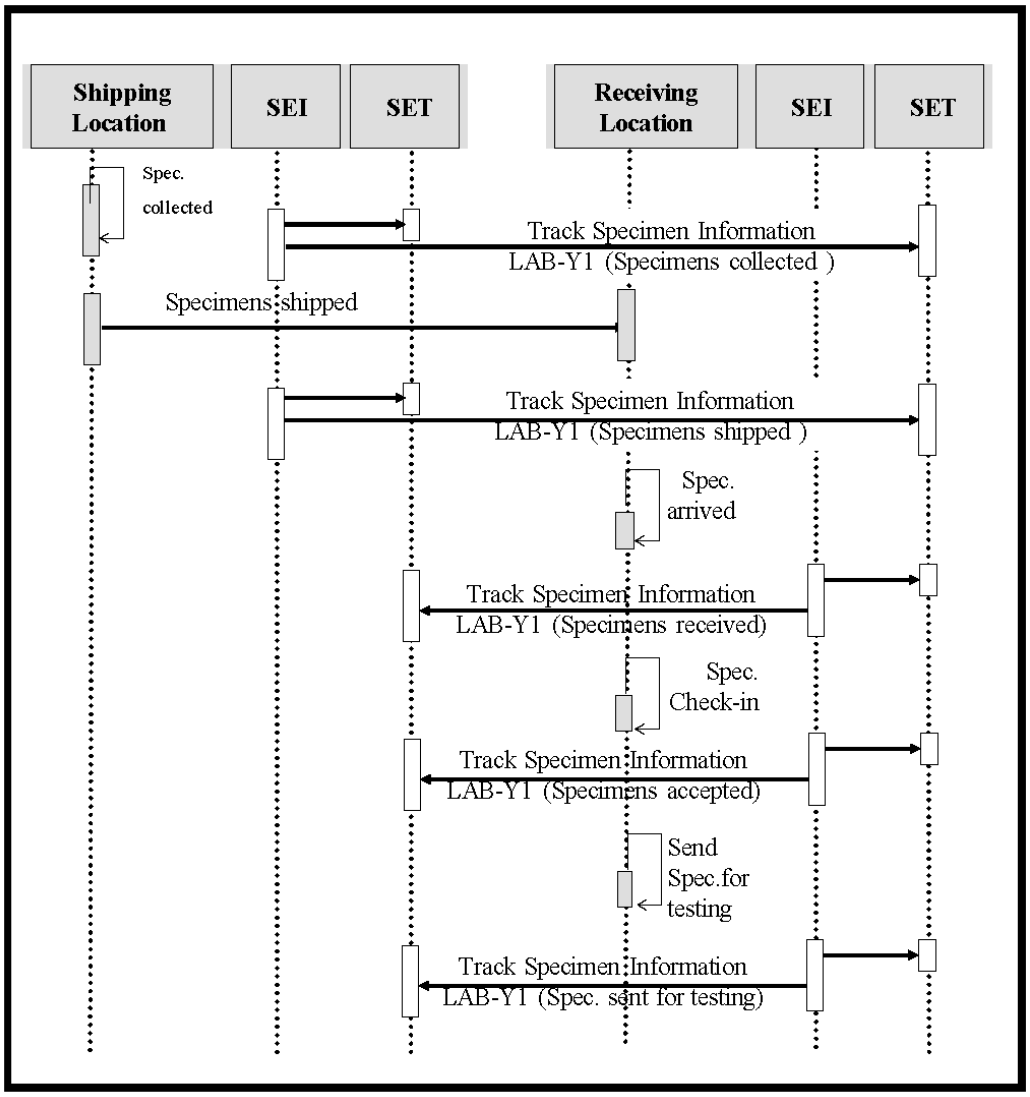

Figure 4.9: **Specimen transferred, no re-identification by receiver process flow.**

4.5.3 Use Case 3: Intra-Laboratory IVD testing specimen tracking

This use case addresses the tracking of the specimen during the overall IVD testing process inside a testing location. The process itself can be more or less automatized, if the testing location is provided with an automation system or not. Only the main generic events are in scope of this profile; in case of automation presence, the atomic tracking of the specimen inside the automation is covered by other profiles (i.e., Laboratory Specimen Handoff). The intra-laboratory IVD specimen tracking use case will focus on these events:

1. The testing manager sends the specimen for testing: it can be a person or a machine, according to the presence of automation;
2. The specimen arrives at a Laboratory Device;
3. The Laboratory Device performs pre-processing, tests execution and post processing operations, according to the needs. Only two events are tracked at this step: processing start and processing end;
4. The specimen is archived or discharged at the end of the IVD devices processing chain.

Steps 2 and 3 could be repeated several times, according to the number of Laboratory Devices the specimen has to visit during the chain itself.

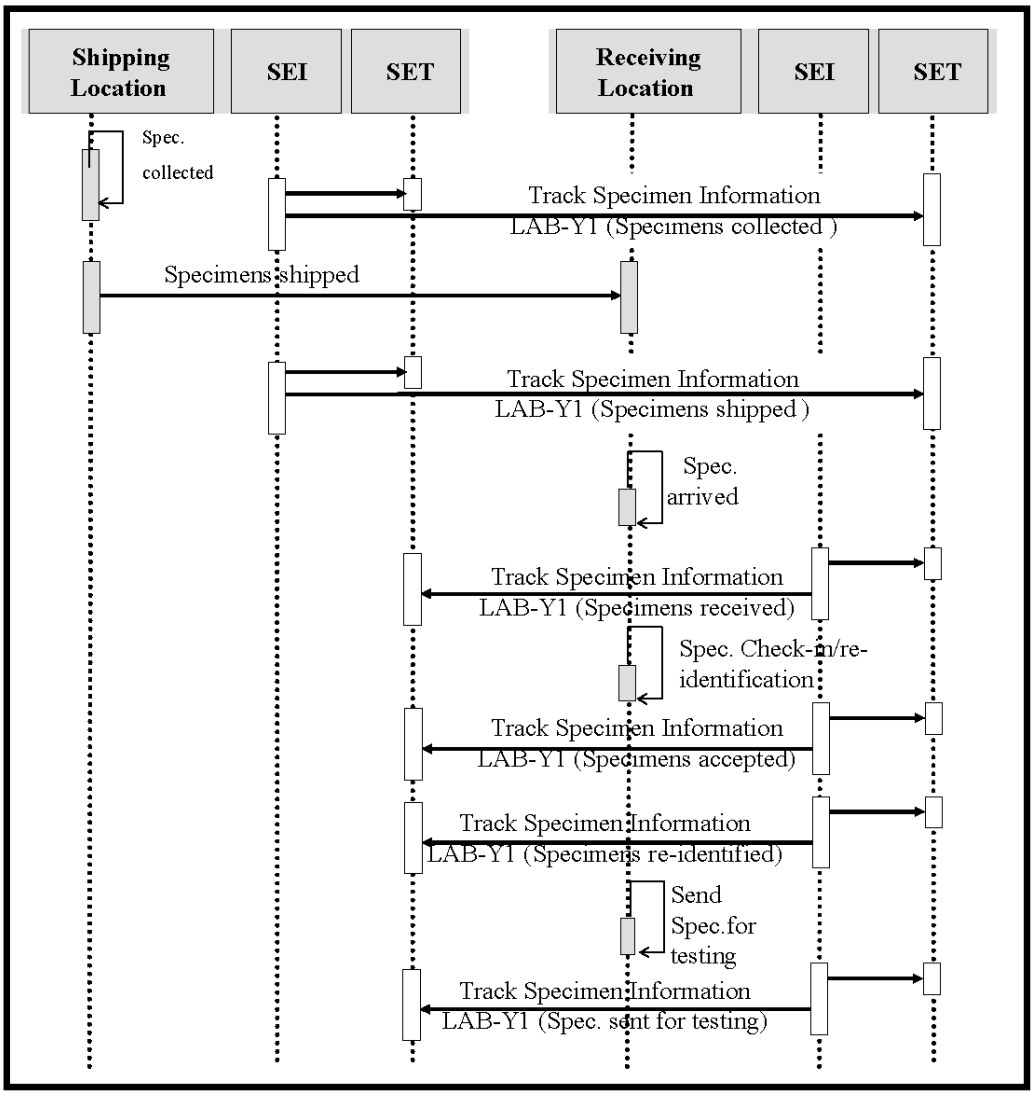

Figure 4.10: **Specimen transferred and re-identified by receiver process flow.**

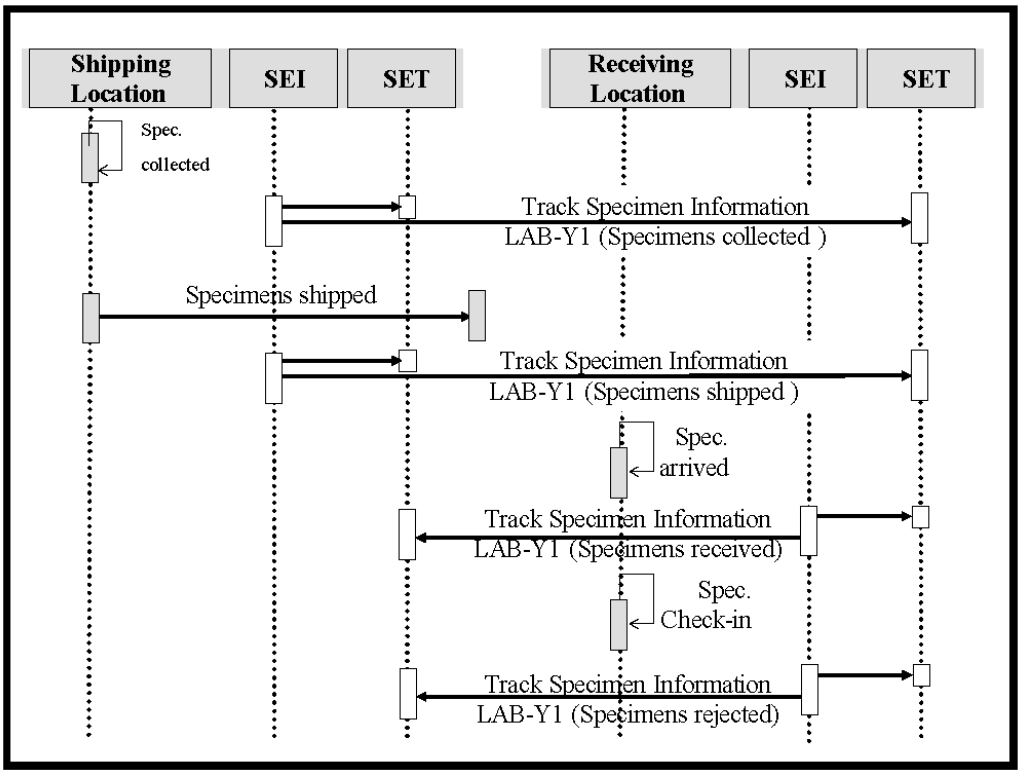

Figure 4.11: **Specimen rejected by receiver process flow.**

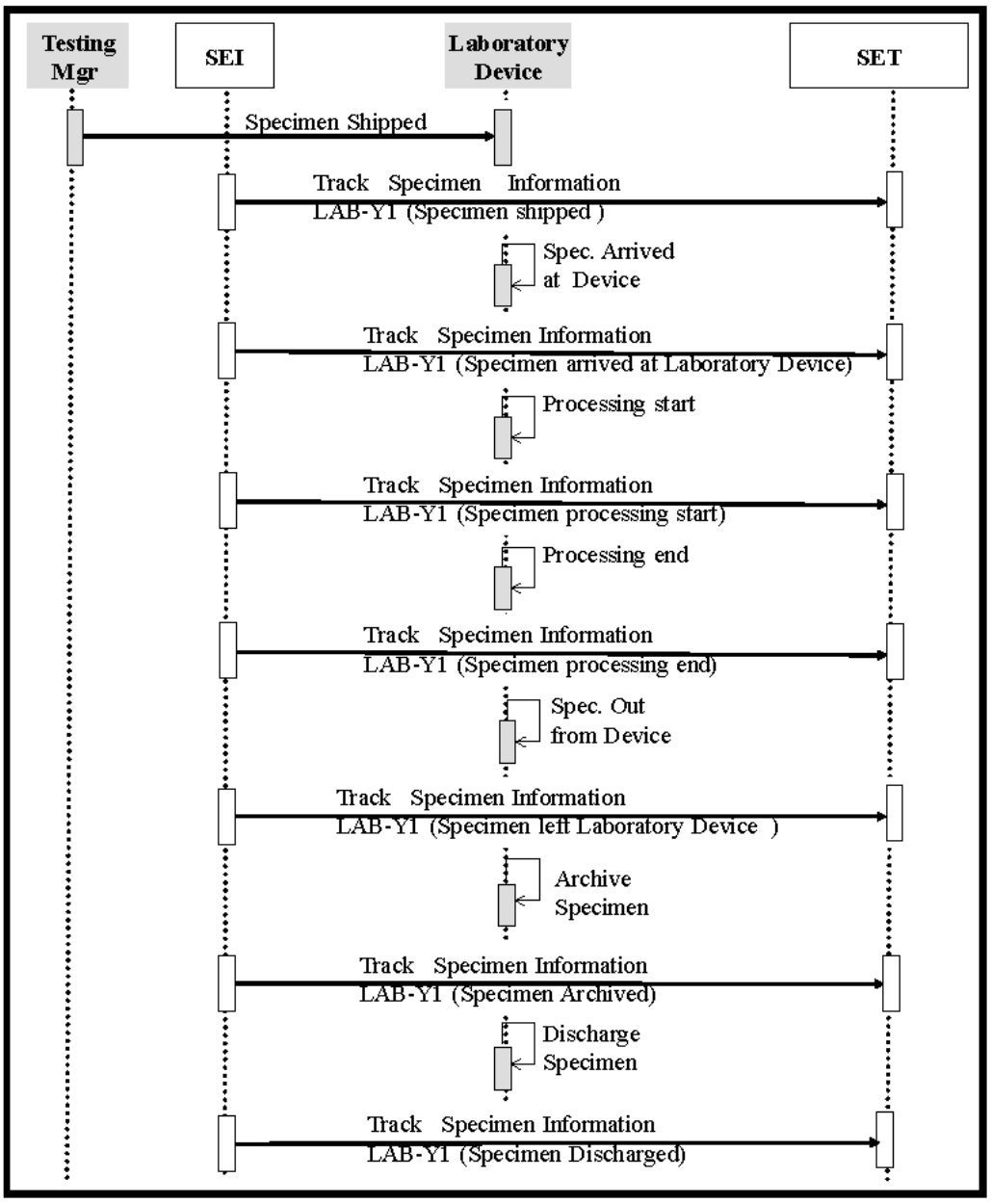

Figure 4.12: **Intra-laboratory IVD testing specimen tracking process flow.**

4.5.4 Use Case 4: Biobank specimen tracking

Biobank workflows and processes vary depending on the research the specific biobank is focused on. Usually biobanks are managed by some reference laboratories, which are hosting them and in most cases belong to university research departments. Several research studies for diseases like cancer, for example, identify a group of volunteers that consent to participate in the research program and to allow collection of one or more samples for the biobank storage. All specimens are shipped to the central laboratory, where technicians process them. At the arrival, the specimen is accepted, identified and archived inside the biobank. When the specimen is required for analysis (selected on query parameters such as class of patients, class of disease, etc.), it is retrieved from the biobank and a new specimen is derived from the original one in order to perform the required tests. This derived specimen is associated to a different container, and then accepted, identified and archived inside the biobank. The original specimen is also archived back inside the biobank, updating some important information as, for example the sample quantity remaining. The SET profile identifies three use cases involving biobanks:

1. Specimen collected in a laboratory and shipped to biobank
2. Specimen retrieved from biobank for immediate testing
3. Specimen retrieved from biobank for testing preparation

There are two main information flows related to the Use Cases above: the first flow is related to sample physical transport, while the second is related to any additional information provided for that sample (clinical_ID, pathology, additional information about the individual and so on). This Use Case is focused only on specimen physical tracking.

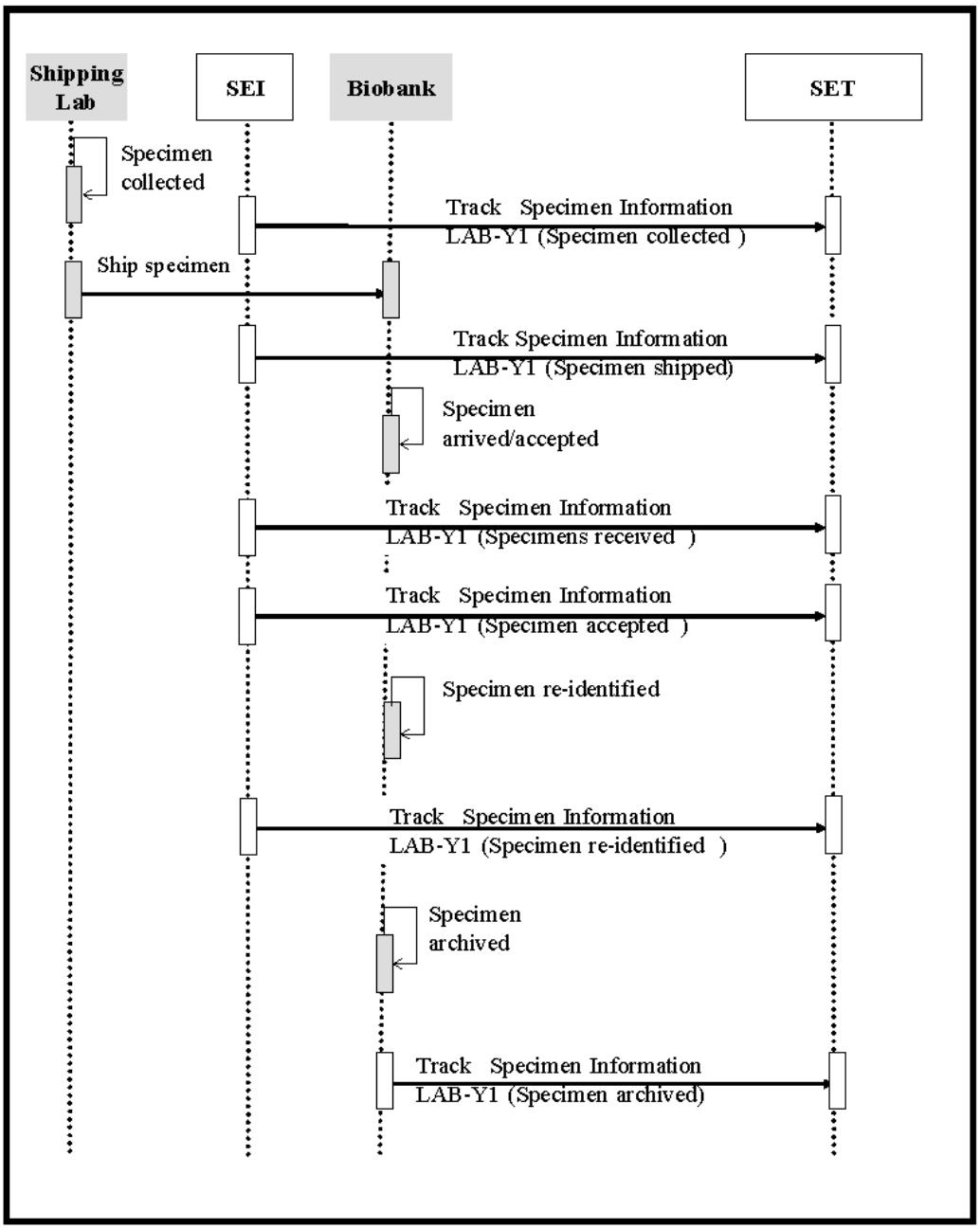

Figure 4.13: **Specimen collected in Laboratory and then shipped to Biobank.**

The use case starts with one of the federated biobank laboratories that collects a specimen for research purposes from a consenting patient. The specimen is collected at one of the federated biobank laboratories, and then shipped to the biobank site, where it is re-identified, according to the biobank information system, and archived for future usage. There are two main cases:

1. The biobank is not allowed to know clinical_ID and clinical data;
2. The biobank is allowed to know clinical_ID and clinical data;

In the first case, sample de-identification occurs before the sample is shipped from the laboratory to the biobank and only the laboratory will keep track of the link between clinical_ID and biobank_ID. When the sample is requested for a trial, the biobank will never be able to provide the clinical information as it hasn't it since the sample arrival in the biobank. In the second case, when the sample is re-identified in the biobank information system, which creates the biobank_ID and assigns it to the sample, both clinical_ID and clinical information are stored in the biobank information system, as metadata associated to the sample. When the sample is requested for a trial, the biobank will provide also the clinical information only if the study protocol allows the use of this kind of information. If the study protocol doesn't allow the use of the clinical information associated to the sample, the biobank information system will have to de-identify the sample, protecting all confidential clinical information. The use case assumes that the specimen is always re-identified in the biobank information system, even if it already has knowledge of the ID previously assigned by the collecting laboratory. Anyway the link between the laboratory assigned ID and the new biobank ID of the specimen is always maintained: in the first case, only by the clinical laboratory, in the second case both by the laboratory and the biobank. From the point of view of the events to track, in both cases we can identify these five below:

1. The specimen is collected at the federated laboratory research site
2. The specimen is shipped to the biobank site
3. The specimen arrives at the biobank site
4. The specimen is accepted and re-identified at the biobank site
5. The specimen is archived for further use

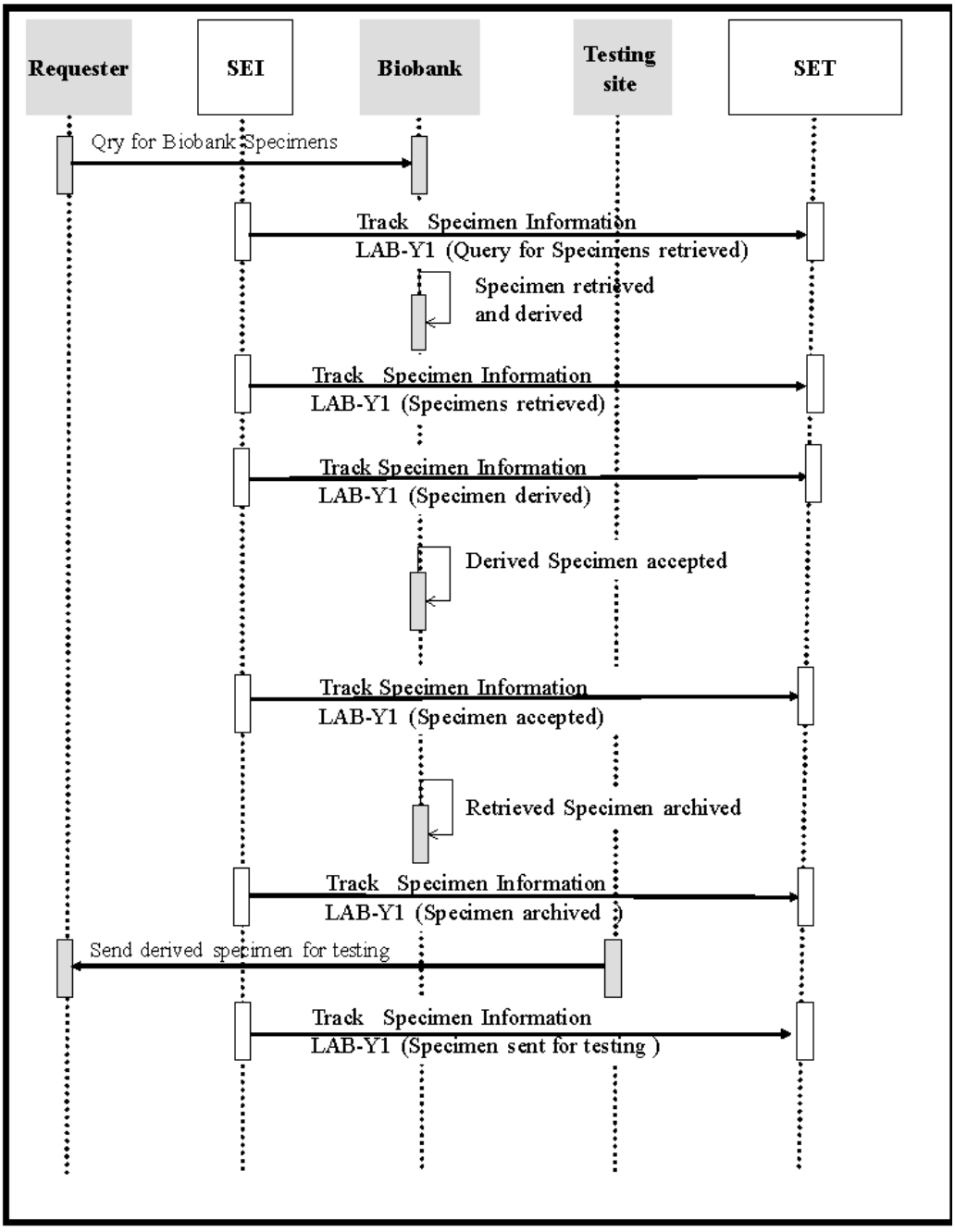

Figure 4.14: **Specimen retrieved from Biobank for immediate testing.**

In this scenario, a query is executed to retrieve one or more specimens from a biobank for immediate testing. Specimens have previously been stored in the biobank with an assigned biobank_ID. If the biobank is not allowed to maintain clinical_ID and clinical data, this information cannot be provided at specimen retrieval. If the biobank is allowed to maintain clinical_ID and clinical data, a set of information regarding patient identifier, name, surname, date of birth, specimen type, related patient disease is also known by the biobank, and can be provided to the requester only if the study protocol allows the use of this kind of information. If the study protocol doesn't allow the use of the clinical information associated to the sample, the biobank information system will have to de-identify the sample, protecting all confidential clinical information. Each of the retrieved specimens is always derived in order to obtain a new specimen that will be used to perform all the required tests: the derivation operation implies the labeling, and consequently the identification, of the derived specimens. The new specimen is immediately identified and accepted in the biobank information system before tests execution. It is very important that the original specimens are archived back to the biobank as soon as possible, updating some important information as the sample remaining quantity. In this use case the derived specimens will be immediately used for testing. After the testing, the specimen can be archived back in the biobank or marked as exhaust if all the sample quantity has been used for testing or if testing itself makes the derived specimen not usable anymore. From the point of view of the SET profile, this use case will track these specimen states:

1. The specimen is retrieved from the biobank
2. The specimen for testing is derived from the retrieved specimen. If the study protocol doesn't allow the use of the clinical information associated with the retrieved specimen, the derived specimens are de-identified
3. The derived specimen is accepted
4. The retrieved specimen is archived again in the biobank. It is very important to update the specimen status: after derivation the specimen can be usable, exhaust, not valid.
5. Tests are performed on the derived specimen

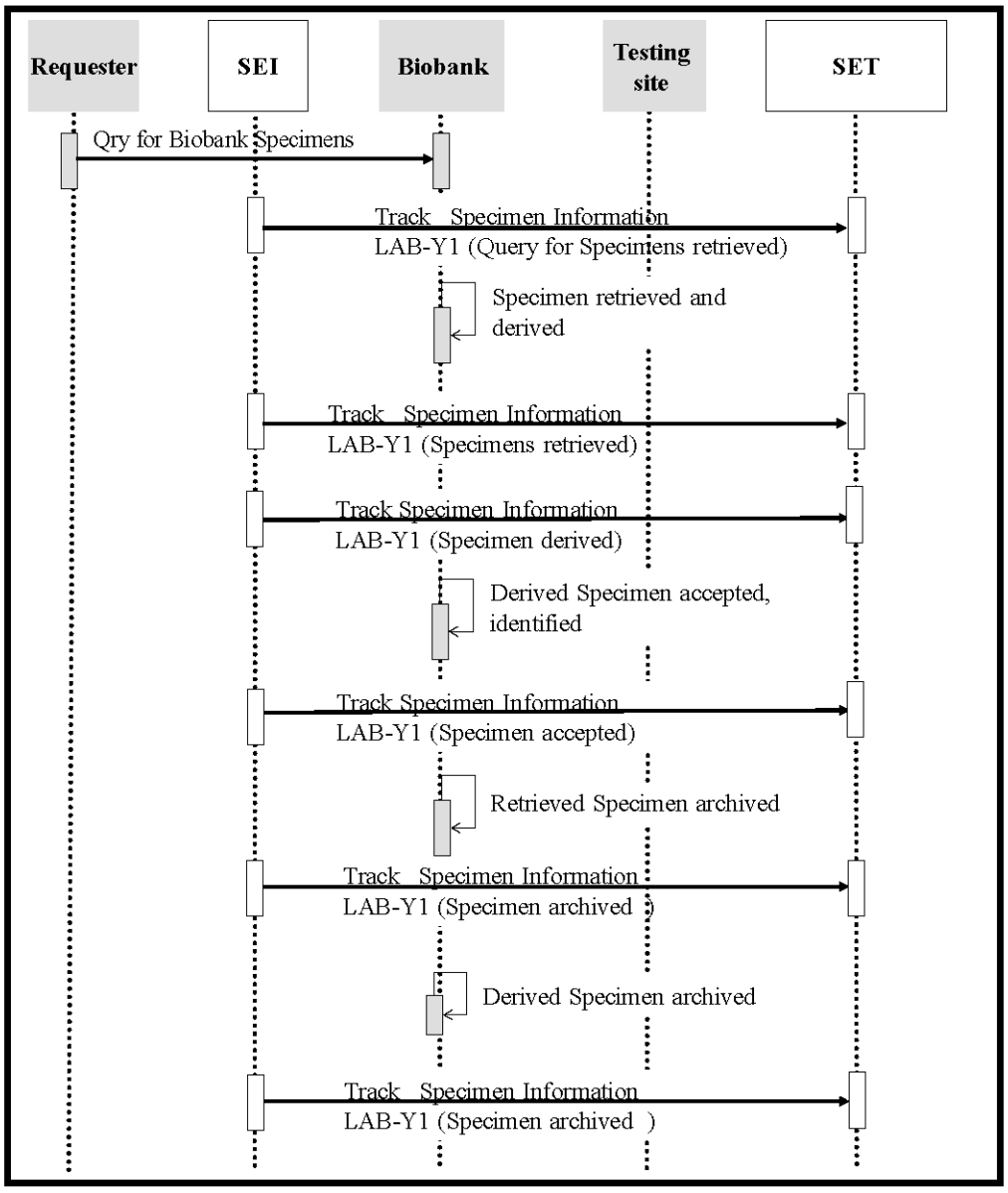

Figure 4.15: **Specimen retrieved from Biobank for testing preparation process flow.**

This use case is similar to the previous one, but the derived specimen is not immediately sent for testing, but it is archived in the biobank for further use. For example such a situation can occur when from a blood sample, retrieved from a biobank, a series of derived DNA extraction specimens are archived to be ready for further testing. As for the previous use case, if the study protocol doesn't allow the use of the clinical information associated with the retrieved specimen, the derived specimens are de-identified.

4.5.5 Use Case 5: Specimen derivation tracking

This use case addresses specimen derivation, which may occur in some Laboratory specialties, like microbiology and pathology. Derivation starts from a "primary" specimen, and through a series of different operations produces one or more derived specimens. Each of them is identified and processed independently from the others. The SET profile related event keeps the link between the primary specimen identifier and the derived ones.

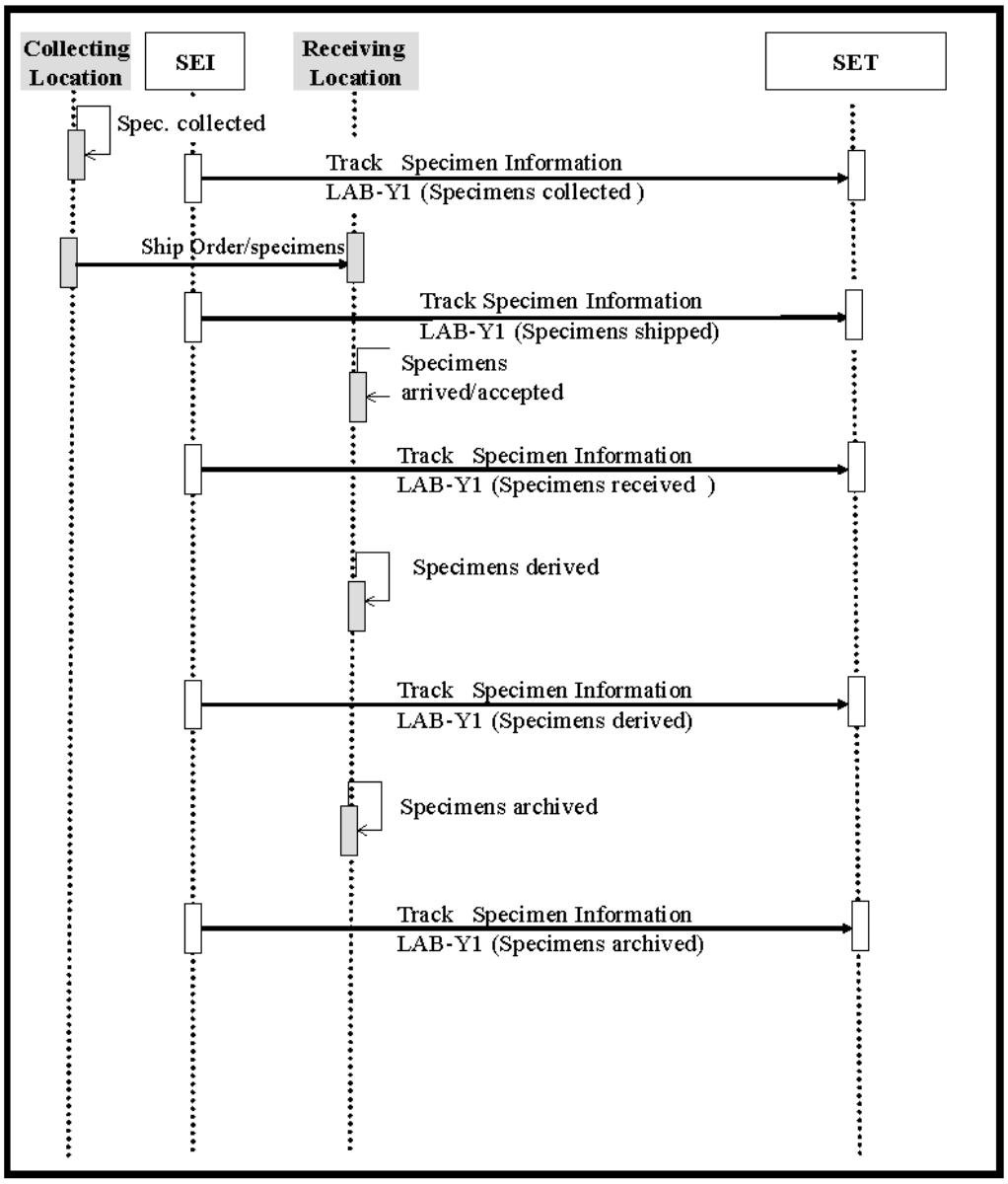

Figure 4.16: **Specimen derivation tracking process flow.**

4.6 Summary and discussion

This Chapter outlined the traceability information model resulting from the work described in this thesis, the Specimen Event Tracking Profile, providing a vision specimen and process oriented for the diagnostic process, according to IHE guidelines. SET aims to support the reconstruction of the entire history of the specimen, from sampling to storage, in a laboratory or in a biobank. The workflow can be recreated thanks to two actors included in the Profile – Specimen Event Informer and Specimen Event Tracker – which exchange traceability data coded in a single transaction, Track Specimen Information. Specimen lifecycle is followed in a series of use cases and modelled buy events. The metadata attached to an event conveyed in the traceability messages leverage the HL7 Specimen Domain Analysis Model. The work which lead to this first version for the proposed profile has been carried on with the extensive collaboration and support of IHE PaLM Technical Committee, as the introduction of this longitudinal perspective, specimen-centered, has been considered interesting by the Committee. The model described is related to what is required by IHE Volume 1 for Technical Framework to create a new profile, that is to say a description of actors and transaction, without specific definition of the messages involved to implement the transaction. The present status of SET Profile includes a series of further activities, out of scope of the original contribution of this thesis, to define the precise standard and structures of the messages. The HL7 v2 messaging standard allows to cover all the SET events with at least three messages, but using different types of HL7 messages for the same IHE transaction is an issue for interoperability. Once published, the SET Profile can also be used to map the events in a diagnostic process for process mining analysis: previous works [94], already shown as IHE profiles can be a valid modelling paradigm for laboratory preanalytical phase and the general vision about the specimen workflow offered by SET can be a formalism suitable to an analysis process oriented of traceability information.

Chapter 5

Introducing process-oriented traceability in openEHR

5.1 Introduction

The openEHR Task Planning model represents an important attempt to combine medical knowledge, task managing and decision support for the operators involved in a care plan. It also opens new research directions in the field of models describing process-oriented traceability, thanks to some of the intrinsic potentialities in these new-born specifications. The Task Planning model is the first formalization for the "process concept" in openEHR specifications: previously, the formalism was focused on the creation of a longitudinal EHR, defining all the necessary building blocks to document clinical care, from the models of medical concepts to all the technical artefacts, interfaces and services to implement it in a robust way. As the work of the last ten years lead to a complete EHR semantic-oriented architecture, more and more adopted and appreciated by clinical institutions and national governments[34], the choice of dealing with the "process dimension" was the natural next step. The design started in 2016, analysing the main workflow formalisms developed in this field - briefly described in Chapter 3 - and the significant experience of Activity Based Design for clinical processes at Intermountain[7]. The vision underpinning openEHR Task Planning Model is summarized in its declaration of intents: "The central concept is that of a plan (or set of plans) designed to achieve a goal and that relates to an active biological subject (normally a human or animal patient), rather than a

passive object, such as a parcel or tissue sample". This sentence also highlights promising directions for further evolutions for the specification, which inspired the work that produced the results described in this Chapter. The trigger element is the fact that Task Planning is mainly focused on the "planning" side, as the name itself suggests, both in the design and in the execution phases. But also, as showed in Chapter 2 and Chapter 3, process-oriented traceability is the opposite side of the "reality coin", for which planning is the first side: therefore, the introduction of process artefacts in openEHR opens new perspectives in modelling traceability too. The work about openEHR specifications described in this thesis resulted in the openEHR Traceability Model, a proposal of complementing the openEHR Task Planning Model. The following paragraphs will outline the original contribution, discussing at the end of the Chapter the emerging critical points and alternative directions.

5.2 The openEHR Traceability model

The openEHR Traceability model proposed in this thesis is a formalism for an openEHR-consistent representation of process-oriented traceability in clinical pathways. The approach followed is based on the idea that planning and traceability are separate domains, with precise scopes and boundaries, but complementing each other's perspective on the natural description of clinical processes. Therefore, the modelling strategy consisted in creating a distinct model, specialized in tracking the dynamics of clinical workflows, as showed in Fig.5.1. The first intent of the openEHR Traceability model is to structure the actual execution path in a process, building the run-time instance of a Task Plan as a bottom-up (from Task to Task Plan) representation of the workflow during care plans. This proposed formalism acts at runtime, tracing all the relevant events and combining the abstract representation of the process with the real behaviour of the activities during execution.

5.3 Main Concepts

5.3.1 Computational Context

The basic idea is that during the execution of a process, performers, applications, sensors and devices create events, which can be collected and used

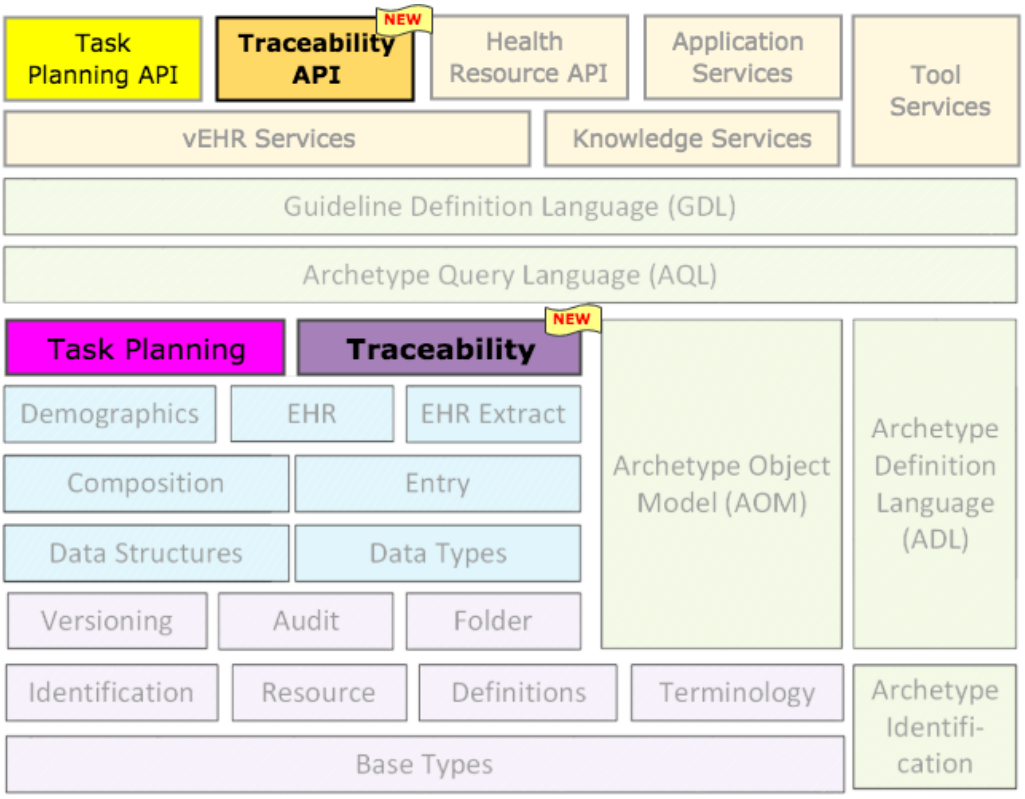

Figure 5.1: **openEHR Traceability Model.** [77]

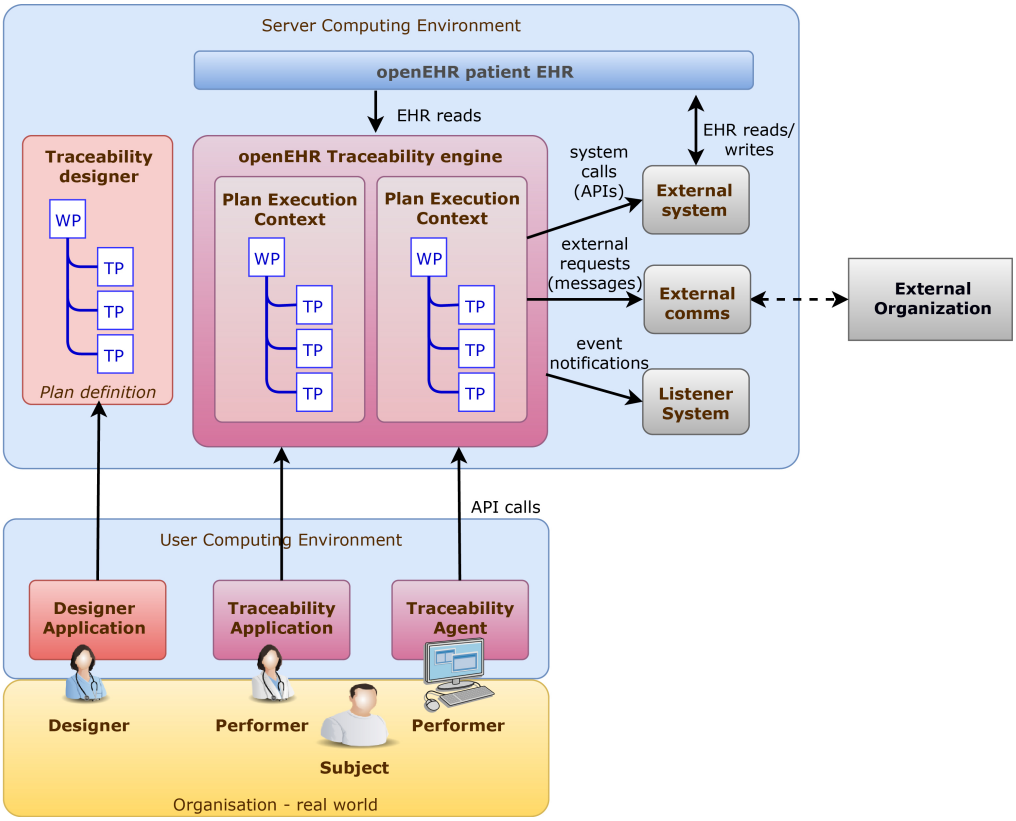

Figure 5.2: **openEHR Traceability Model Computational Environment.** During the execution, the Traceability Engine interacts with the actors in the computational environment in analogy with the Task Planning Engine described in openEHR Task Planning specifications[30]

to reconstruct the control flow thanks to the information artefacts defined in openEHR Task Planning and the openEHR Traceability model, developed in this work. In analogy with the Task Planning Engine, this proposed specification describes the models and semantics for traceability by a notional openEHR Traceability execution engine ('Teng'). Teng operates in a computing environment including other systems and can communicate with them. A key system is the openEHR EHR system, via which openEHR patient EHRs are accessible to the Traceability Engine. Outside the environment there are other organisations with which Teng engine interacts via an appropriate communication/notification system within the environment. The basic operational mode consists of a series of Traceability Agents, which can be simple modules associated to sensors generating events or more sophisticate applications, interacting with the Task Principal Performer and

giving feedbacks related to the ongoing process. Traceability Agent communicate with the Teng via its API and can be implemented as applications with which the Users can interacts or via services running independently and tracing part of the process - for example via sensors non directly managed by Users - when Tasks are performed by devices or software agents. These agents may be stand-alone or combined with Task Planning applications or other applications.

The overall context is in Fig.5.2.

5.3.2 Artefacts

The central artefacts for the model are two classes (`TRACE_EXECUTION_HISTORY` and `TRACE`) dedicated to support the process execution path reconstruction from a series of events.

The other proposed classes are the dual version of some of the Materialise model classes: as materialised classes are the artefacts to actualise the speculative process during plan execution, the "tracing" classes are the artefacts to detect the process dynamic during or at the end of the execution. The classes are listed below and will be outlined with more detail in next paragraphs:

- `TRACE`: is the footprint of the Traced Work Plan, a strictly chronological picture of the recorded Traced Task.
- `TRACE_EXECUTION_HISTORY`: history of Traced TaskPlan execution events and notifications, in chronological order.
- `T_TASK`: tracked version of the abstract Task, instantiated when the first Event Record related to the Task is received by the Traceability Engine.
- `T_ACTION`: tracked version of the abstract Action, created and updated as the Task evolves in its lifecycle.
- `T_TASK_PLAN` and `T_TASK_GROUP`: tracked version of the abstract Task Plan and Task Group, instantiated to group the Traced Tasks present in a Trace.
- `T_WORK_PLAN`: tracked version of the abstract Work Plan, instantiated to group the Traced Task Plan and Traced Task Group present in a Trace.

In the tracking strategy, also the Definition and History model artefacts are used, at design time and to model the Events defined to trace the process execution.

5.3.3 Tracking Strategy

The openEHR Tracebility model can be implemented in different conditions and is compatible to computing environment with or without Task planning support during Tasks' execution. The operation mode is summarized in the following steps:

- At design time, an abstract Work Plan and its constituent Task Plans are created to model each process or sub-process that will be analysed from a traceability perspective. In this phase, the Task Plans are represented in the form of openEHR templated archetypes based on the Task Planning definition model.

- When execution starts, the Work Plan is instantiated and is used as a concrete definition of the Task Plan for a specific subject in a specific situation. An EXECUTION HISTORY is also created for each Task Plan.

- At the beginning of the real world process, for each occurrence related to the execution and traced by the Traceability Agents, the Traceability Agents generate Events, which are modelled as instances of the `*_EVENT_RECORD` classes defined in these specifications and in the Task Planning model.

- The Events are sent by the Traceability Agents to the Traceability Engine: each Event contains a `Task_Id` and a series of structured information about the process behaviour.

- As the first Event for a Task is received by the Traceability Engine, the Teng creates the correspondent `T_TASK`, instantiates a TRACE object and adds to the TRACE object the `T_TASK` created.

- As the process advances, the Traceability Agents distributed along the process will create Events, triggered by workflow steps or external notifications.

- If the Event received is associated to an already instantiated T_TASK, the Traceability Engine will save the Event in the TRACE_EXECUTION_HISTORY and the T_TASK will be updated in the TRACE object.
- If the Event received is not associated to an already instantiated T_TASK, the Traceability Engine will save it in the TRACE_EXECUTION_HISTORY, will create the T_TASK object and will add it to the TRACE object.
- During execution or at its end, the Traceability Engine also groups the T_TASKS contained in the TRACE object, instantiating the T_WORK_PLAN object and creating the T_TASK_GROUP/T_TASK_PLAN objects defined by the abstract Work Plan and containing the model for the T_TASK effectively executed.

A Trace object is a list of the T_TASK ordered by Task's event timestamps. From a traceability point of view, a single TRACE can be analysed to obtain information about process execution in the real world, or it can be lead back to the underlying model in terms of T_TASK_PLAN, derived from the correspondent Task Plan created at design time. The logic of passing from a TRACE to a TASK_PLAN is implemented and managed by the Traceability Engine, according to the execution context, the model defined in the planning phase and to specific organizational needs. For each Task it is also possible to reconstruct the effective lifecycle. The Traceability Application can also rely on the Execution History, which records all the events related to the execution path for each Task Plan and can be used to support the analysis of events, an example being cases in which Events not directly associated to a Task are created. Traced Task Plan can be created and showed iteratively during process execution for operational support, but also when execution is completed, for a post-mortem analysis.

5.3.4 Time

The Traceability model works in the world of "what has happened", therefore the distinction present in the Task Planning model - where time can be a relative offset, an absolute time or the time of an event – in no longer present. All the Traceability Agents are synchronized, with a time precision configured at millisecond level. The possible delay between the time in which the Event happens and is recorded by the Traceability Agent and its notification is sent to the Traceability Engine is not an issue, as Tasks are ordered in Traces according to their timestamp.

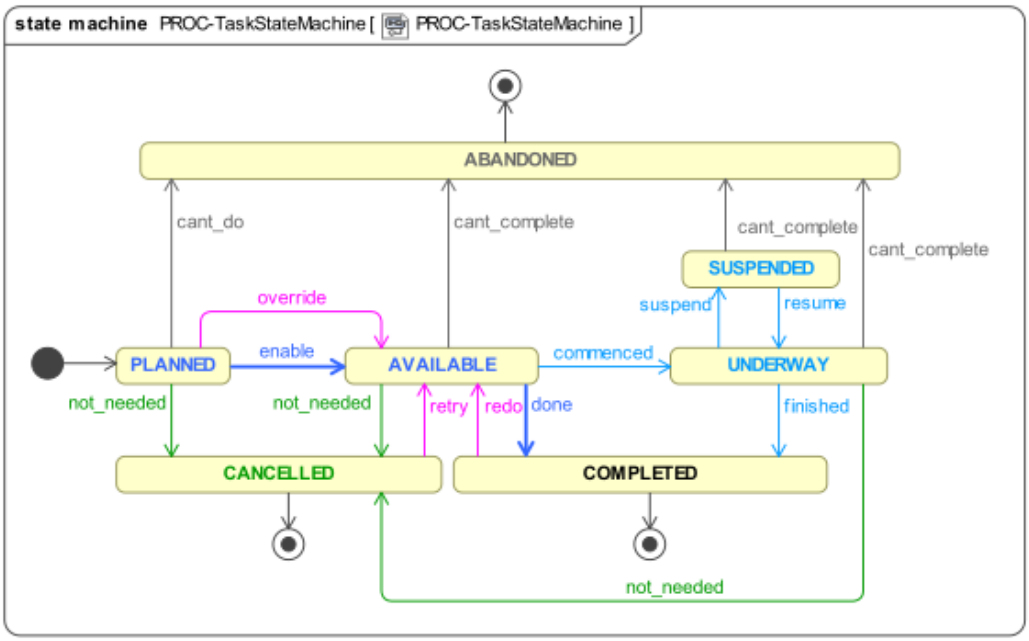

Figure 5.3: **Task State Machine.** At each step of the task lifecycle described in [30], Events can be emitted to track the Plan execution

5.3.5 Events

The events considered in the Traceability Model are those defined in Task Panning Model, that is to say:

- timer event: an event caused by a timer that expires;
- state trigger: a condition based on one or more observed variables available from the computation environment, including subject variables (sex, vital signs, etc) and variables relating to clinical processes, e.g. 'time since emergency admission';
- task transition: an event generated by the state transition of a Task during execution so far, e.g. the previous Task completing;
- callback notification: an event generated by a notification received on completion of a Task dispatched to a different Plan or an external system;
- system notification: an external event notified to the system by a user, e.g. receiving a phone call.

- manual notification: an external event signalled manually to the system by a user, e.g. receiving a phone call

The Events related to the task transitions can be configured following the model defined by the Task State Machine in Fig.5.3, which structures the task lifecycle.

All the traced Events are saved in the `TRACE_EVENT_HISTORY` by the Traceability Engine in chronological order and used to monitor and save the Traced Task chain.

5.3.6 Clinical investigator process

Tracing can be extended to all the steps of a cognitive loop of care, creating and configuring a series of Traceability Agent to detect the phase behaviour, as showed in Fig.5.4. Abstract model definition, via the Task Planning model formalism, is the starting point both for runtime task execution support and traceability support, they can coexist but they are independent. In a real implementation, in fact, the Traceability Engine can be present both with or without the Task Planning Engine, and when they are contemporarily present they can be implemented in the same application or in two separate systems.

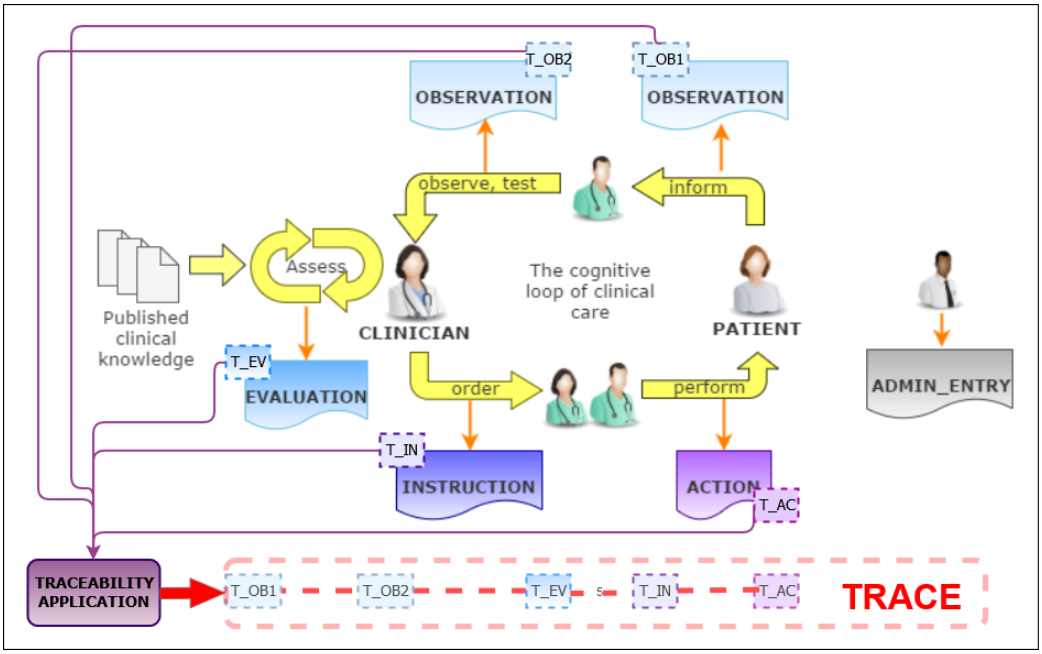

Figure 5.4: **Traceability in the clinical investigator process.** The figure shows how Traced Tasks (`T_TASK`) of various kinds are collected in a TRACE object during or after execution.

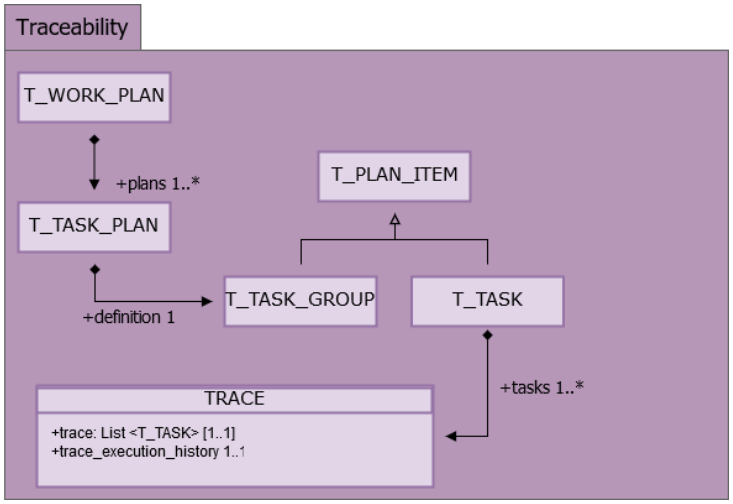

Figure 5.5: **Traceability package.**

5.3.7 Package and classes

Fig.5.5 shows the proposed Traceability package. This is a partial specification of classes that could be instantiated during the execution of a process, to record Tasks and System Events and to create a structured representation of the Traced Work Plan.

The central classes of the package are the `TRACE` and `TRACE_EXECUTION_HISTORY` classes, as they contain the basic artefact to create a "process footprint", also usable for process mining applications. The other classes are the Traced version of the corresponding classes in the Materialisation model.

Class	TRACE	
Description	The chain of Process instances in the form of a list of T_TASK, instantiated and updated as new Events are recorded. It is used by Traceability Engine to record the chronological order of Traced Tasks during execution, according to the timestamp of the first event recorded for a specific Traced Task. May be persisted in the EHR in partial or complete form, or not at all.	
Attributes	**Signature**	**Meaning**
1..1	**trace**: List<T_TASK>	Traced Tasks instantiated and updated as Events arrive.
1..1	**trace_execution_history:** TRACE_EXECUTION_HISTORY	History of Traced Plan execution events and notifications.

Figure 5.6: **Trace Class.**

Class	TRACE_EXECUTION_HISTORY	
Description	History of Traced Plan execution events and notifications. It is used by Traceability Engine to record all events during execution, in a chronological order. May be persisted in the EHR in partial or complete form, or not at all.	
Inherit	TASK_PLAN_EXECUTION_HISTORY	
Attributes	**Signature**	**Meaning**
0..1	**notification_events**: List<TASK_PLAN_EVENT_RECORD>	List of notifications to another party (i.e. performer).

Figure 5.7: **Trace Execution History Class.**

Class	T_WORK_PLAN	
Description	Traced form of WORK_PLAN.	
Attributes	**Signature**	**Meaning**
0..1	**plans**: List<T_TASK_PLAN>	Member Traced Plans.
1..1	**owner**: WORK_PLAN	Owning Work Plan definition.
1..1	**timeline**: T_TIMELINE	Tracked global timeline.
1..1	**context**: T_PLAN_DATA_CONTEXT	Traced global context.
0..1	**materialised_workplan**: List<LOCATABLE_REF>	Forward references to all the Work Plans materialised from the owning Work Plan definition.

Figure 5.8: **Traced Work Plan Class.**

Class	T_TASK_PLAN	
Description	Root object of a traced Task Plan structure, used in execution or at the end.	
Attributes	**Signature**	**Meaning**
0..1	**start_time**: Date_time	Nominal start time for the Task Plan as a whole. The timings of individual Traced Tasks are absolute.
1..1	**definition**: T_TASK_GROUP	Root of Task Plan Task structure.
1..1	**owner**: TASK_PLAN	Owning Task Plan definition.
1..1	**principal_performer**: T_PERFORMER_ALLOCATION	Run-time principal performer - a person or other agent.
Functions	**Signature**	**Meaning**
	lifecycle_state (): T_TASK_LIFECYCLE	Lifecycle state of Task Plan, derived as a copy of the *lifecycle_state()* of the definition TASK_GROUP.

Figure 5.9: **Traced Task Plan Class.**

Class	*T_PLAN_ITEM (abstract)*	
Description	Abstract parent of traced run-time types that correspond to design time PLAN_ITEM instances.	
Attributes	**Signature**	**Meaning**
1..1	**definition**: PLAN_ITEM	Corresponding item from Task Plan definition. Redefined in descendants to the definition type corresponding to each runtime (RT_XXX) type.

Figure 5.10: **Traced Plan Item Class.**

Class	T_TASK_GROUP	
Description	Traced form of TASK_GROUP from a Task Plan definition.	
Inherit	T_PLAN_ITEM	
Attributes	**Signature**	**Meaning**
0..1	**members**: List<T_PLAN_ITEM>	Member run-time items in a group, according to the structure of the corresponding definition group instance.
1..1 (redefined)	**definition**: TASK_GROUP	Reference to corresponding TASK_GROUP instance in Task Plan definition.
Functions	**Signature**	**Meaning**
	lifecycle_state (): TASK_LIFECYCLE	Effective lifecycle state, computed from the states of members of the group.

Figure 5.11: **Traced Task Group Class.**

Class	T_TASK	
Description	Traced form of TASK type and its descendants from a Task Plan definition.	
Inherit	T_PLAN_ITEM	
Attributes	**Signature**	**Meaning**
1..1	**lifecycle_state**: TASK_LIFECYCLE	Current lifecycle state of this Traced Task at run-time.
1..1	**task_events**: List<TASK_EVENT_RECORD>	List of references to run-time Task Events that have occurred on this Task.
1..1	**trace_events**: List<TRACE_EVENT_RECORD>	List of references to run-time Trace Events that have occurred on this Task.
1..1 (redefined)	**definition**: TASK	Reference to corresponding TASK instance in Task Plan definition.

Figure 5.12: **Traced Task Class.**

Class	T_TASK_ACTION	
Description	Traced form of TASK_ACTION type.	
Attributes	**Signature**	**Meaning**
1..1	**definition**: TASK_ACTION	

Figure 5.13: **Traced Task Action Class.**

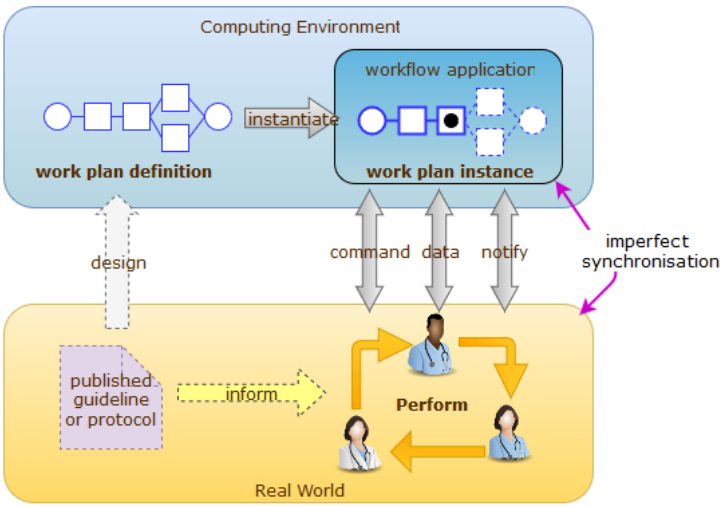

Figure 5.14: **Work plan execution paradigm.[30]**

5.4 Relationship with Task Planning and Entry model

The Traceability model proposed can be seen as a dual form of the Materialisation model defined in openEHR Task Planning specification. The distinction between definition and execution for a Plan makes openEHR Task Planning more flexible and suitable to model clinical processes, which are strongly adaptive, declarative and non-deterministic. The Tasks are semantically defined and directly interpretable by an execution engine, which can communicate with the real world via commands, data and notifications. Anyway, despite this attempt to manage the complexity of reality, there will still be two executing workflows, one virtual and one real, whose synchronisation is imperfect, as the openEHR specification itself depicts in Fig.5.14.

The Traceability model proposed in this work is an attempt to reduce the gap between plans and reality as it tries to create a faithful picture of what happened during process execution. In Task Planning specifications, the History model has a similar objective, but considers a non-structured approach. The Traceability model scope is therefore a complement to the objectives of the Task Planning model. Traced Task can be associated to the Entry objects created during the execution of the activities associated with the Tasks. The Task Planning specifications presented an issue[1] about the usefulness of

[1]ISSUE-fwd-refs: entry_instance is a forward reference - which requires updating the

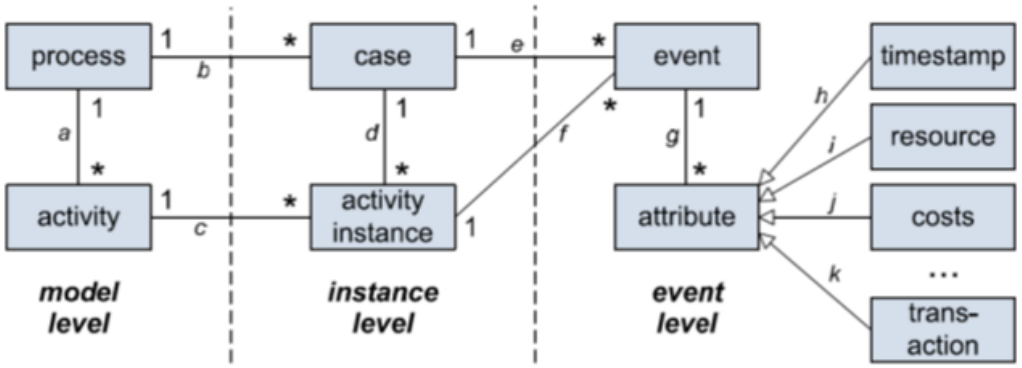

Figure 5.15: **Process Mining Typical Scenario.[3]**

forward-reference to the relevant Entries generated by Tasks execution, but for traceability applications this issue is mitigated as being able to determine the resulting Entries starting from the Traced Task Plan is surely useful. The exact implementation strategy (add the links to the Entries or querying to find them from the Plan items) can probably be determined during Traceability Engine implementation, as the decision can depend upon the specific use.

5.5 Process mining applications

The tracking strategy previously outlined considers a two-step approach to detect the execution path: firstly, structured information deriving from runtime events is used to represent what is happening, instantiating a series of Traced Tasks, modelled following abstract Traced Task Planning specifications. Secondly, the execution Trace is used as a basis for reconstructing the Traced Work Plan, according to the conceptual Work Plan instantiated for the process. The defined Work Plan is therefore used to remap the Traced Tasks, chronologically saved in the Trace object, in the form of a series of Task Group/Task Plans. On one side, the second step of reconstruction has the added value of evolving the traced process representation, passing from a one-dimensional to a bi-dimensional representation of the execution path. On the other side, this reconstruction is biased by the "a priori" knowledge

Task Plan after the Tasks have been performed and relevant Entries committed. Is this complication worth the benefit obtained, i.e. directly followable links rather than querying, to find Entries from the Task Plan items? Is being able to determine the resulting Entries starting from the Task Plan even useful?

used to read the information contained in the Trace. To improve the reduction of the gap described in the previous Paragraph, between the computing environment and the real world, the tracking strategy proposed can also be used also as a support for process mining analysis, with the methodologies briefly introduced in Chapter 3. The typical scenario Fig.5.15 for any process mining analysis requires the presence of an event log connected to an initial process model, which may have been constructed manually or discovered through previous process mining steps. As shown in the following figure, openEHR archetypes and classes can be used to model and support each step of process mining, in particular:

- the model level can be implemented using the Task Planning Definition Model to create the speculative process model in terms of Work Plan and Task Plans and the activities in terms of Tasks;
- the instance level can be implemented using the artefacts of the Traceability model, representing with a Trace the execution path described by a case (which is a process instance) and the executing/executed activity with a Traced Task;
- the event level can be implemented using Event Record objects to model the events, as all the more relevant attributes to reconstruct the executing process are present in the classes `TASK_EVENT_RECORD` or `TRACK_EVENT_RECORD`.

More specifically the association could be:

- Process - Work Plan
- Activity - Task
- Case - Trace
- Activity instance - Traced Task
- Event - Task and Plan Event Record
- Attribute - Event Record's classes attribute

A traceability analysis based on process mining techniques, starting from the detection of a series of Traces and Traced Event Records, would create a discovered process model that could be used to better understand the real

world evolution of the clinical process and compared to the original Work Plan to improve the theoretical model representation. Also other kind of events or attributes could be added, according to the perspective that is considered in the analysis. The adoption of openEHR for modelling process mining artefacts also ensures a mitigation to data quality issues. Data quality is a transversal theme for all data-driven analysis systems in general, as the analysis potentiality and results strictly depend on the quality of the available information, from a syntactic and semantic point of view. Fig.5.16 shows how openEHR modelling for data and processes, also including the proposed Traceability model, can successfully address the G4L, 12 guidelines for logging described in [3].

5.6 Summary and discussion

In this Chapter a proposal for an openEHR model devoted to capture the dynamics of clinical processes is described: the openEHR Traceability model. The model benefits of the potentiality about process modelling introduced by the Task Planning specifications, exploring one of their possible directions of application and evolution, the traceability perspective. The basic idea is to model the process to be traced in terms of abstract Work Plans and to build its traced representation, the Trace, creating and updating Traced Task as the associated Events are sent by a series of Traceability Agents. The Tasks are saved in the Trace chronologically and then grouped in a Traced Workplan, following the logic of the model designed. The main conceptual elements of the Traceability model have been described, including basic artefacts, computational environment and relationship with other openEHR models. A strategy to improve the model with process mining techniques has also been described, highlighting the advantages of using openEHR to model process and event logs for process mining analysis. The first discussion point has to be dedicated to the opportunity of creating a separate model in openEHR instead of a package, similar to the Materialisation model package, but focused on traceability. This question emerged in several moments of the work described in this thesis, as the previously cited duality also opens the way to include traceability-oriented artefacts in a package included in the Task Planning model. The choice of creating two separate models derives from considerations about the objective and the scope of the Task Planning model. The specification, as the name itself indicates, has been created with the intention of supporting clinicians during patient care from the planning of the activities to their execution. Future and present are therefore the pre-

Logging Guideline	How the Guideline can be addressed by openEHR approach
[G4L1] Reference and attribute names should have clear semantics.	All openEHR artifacts have a clear semantics, it is one of the formalism's strengths
[G4L2] There should be a structured and managed collection of reference and attribute names.	OpenEHR specification are a structured reference and have precise governance rules to manage their development
[G4L3] References should be stable.	Specifications, after development phase, are available in a stable version
[G4L4] Attribute values should be as precise as possible.	The representation of attributes with openEHR archetypes support an high level of detail, which can be further extended
[G4L5] Uncertainty with respect to the occurrence of the event or its references or attributes should be indicated immediately	The uncertainty recording depends on the logging application, but its value can be included in the Task Event Record class
[G4L6] Events should be at least partially ordered.	Trace Event Record, for each instance, contain the list of the logs chronologically ordered
[G4L7] If possible, also store transactional information about the event.	Task Event Record registers the Traced Task lifecycle state when the event is generated.
[G4L8] Perform regularly automated consistency and correctness checks. Ensure provenance.	Consistency and correctness checks can be implemented in the Traceability Engine application. A provenance system can rely on openEHR versioning and auditing support.
[G4L9] Ensure comparability of event logs over time and different groups of cases or process variants.	The use of openEHR ensures long-term comparability of event logs over time and over process variability
[G4L10] Do not aggregate events in the event log used as input for the analysis process.	This is related to the tracking application logic
[G4L11] Do not remove events and ensure provenance.	Depending on the tracking application logic
[G4L12] Ensure privacy without losing meaningful correlations.	Supported by openEHR separation between demographics and clinical information

Figure 5.16: **Work plan execution paradigm.[30]**

dominant dimensions; also the past is taken in account, but has a minor role. For these reasons, a separate model is proposed, to clearly divide objectives and scopes. Another point is related to the timeline of the development of Task Planning specifications: they are recent, they will be soon evaluated in a real context but at the moment there is no reference implementation. Nevertheless, the youth of this model can be also viewed as a good moment to express a proposal, as the relevant element could be inspiring also for the further development of Task Planning. In this sense, the sophisticated generalisation reached in the model formalization allows to also model plans related to specimens, extending the concept of "active biological subject". In fact, adding to the Task description the archetypes modelling specimen information, it would be possible to easily include the use cases supporting the SET Profile, described in Chapter 4. Probably lower level of logging, like the work orders associated to analysers in a laboratory, could be out of openEHR scope, but but a specimen-focused activity support at a high level could be considered for planning, support and tracking aims.

Chapter 6

Conclusion

This thesis examines how science and technology can, at the same time, support and be stimulated by the growing interest in traceability expressed by the population, institutions, industry and research. In medicine, the attention to traceability is particularly felt, because it is vital – and, in most cases, not in the figurative sense - that the level of safety, reliability and quality of devices, data and care pathways is adequate and clear at the time of use. Nevertheless, as the capacity offered by technology increases, so do the difficulties in using the heterogeneous data generated by the plethora of tools and methodologies that are gradually being created. The work presented in this thesis addresses these issues proposing two information models describing clinical process-oriented traceability in IHE and openEHR specifications. The first result, the IHE Specimen Event Tracking (SET) Profile, models in IHE guidelines the traceability for a biological sample in a diagnostic process, covering specimen life cycle from collection to storage or discard, in clinical and research contexts. The second result, the openEHR Traceability Model, defines traceability as the complement of planning, recently introduced in openEHR specifications by the Task Planning Model specifications.

All the results are theoretical, therefore the following steps are dedicated to consolidating them and completing what is still missing to arrive at a reference implementation in one or more selected use case, which could also be designed to combine the two models in a unique prototypal application. The SET Profile result is in a more advanced state, as it is one of the current projects considered in IHE PaLM Committee development timeline. The part of SET realized in this work is related to the Volume 1 of the Technical Framework, focused on the Profile description, in terms of Actors, Trans-

actions (described at a high level) and Use Cases. The detailed definition, always in collaboration with the IHE PaLM Committee, of SET Transaction is the first of the future developments, already in progress at this time, and include the creation of new HL7v2 messages, which will be proposed to the HL7 organization to better capture process-oriented traceability for specimens. The openEHR Traceability Model result is in a more preliminary step as it has still to be presented to the openEHR Specification Program. Future works will therefore start with the contacts with openEHR to propose the model presented in this dissertation and will evolve according to received suggestions.

Bibliography

[1] ISO 17511: 2003. *In vitro diagnostic medical devices: measurement of quantities in biological samples: metrological traceability of values assigned to calibrators and control materials.* (Cit. on p. 10).

[2] AACC. *American Association for Clinical Chemistry*. 2017. URL: https://www.aacc.org/ (visited on 01/2019) (cit. on p. 10).

[3] W.M.P. van der Aalst. *Process Mining: Data Science in Action*. Springer Berlin Heidelberg, 2016. ISBN: 9783662498514. URL: https://books.google.it/books?id=hUEGDAAAQBAJ (cit. on pp. 16, 23–26, 75, 77).

[4] Wil M. P. van der Aalst. "Business Process Management Demystified: A Tutorial on Models, Systems and Standards for Workflow Management". In: *Lectures on Concurrency and Petri Nets: Advances in Petri Nets*. Ed. by Wolfgang Desel Jörg and Reisig and Grzegorz Rozenberg. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 1–65. ISBN: 978-3-540-27755-2. DOI: 10.1007/978-3-540-27755-2_1. URL: https://doi.org/10.1007/978-3-540-27755-2_1 (cit. on p. 14).

[5] Dave Armbruster and James Donnelly. "Harmonization of clinical laboratory test results: the role of the IVD industry". In: *EJIFCC* 27.1 (2016), p. 37 (cit. on p. 10).

[6] Food Standards - Australia and New Zeland. *Food traceability*. 2017. URL: http://www.foodstandards.gov.au/industry/safetystandards/traceability/pages/default.aspx (visited on 01/2019) (cit. on p. 5).

[7] T. Beale. *Initial foundations for clinical workflow*. 2017. URL: https://wolandscat.net/2017/01/23/initial-foundations-for-clinical-workflow/ (visited on 01/2019) (cit. on p. 59).

[8] International Society for Biological and nvironmental Tepositories. *Standard PREanalytical Code (SPREC) v2.0*. URL: https://www.isber.org/page/SPREC (visited on 01/2019) (cit. on p. 11).

[9] Piero Cappelletti. "Brain-to-brain loop 2020: è ancora utile il ciclo di Lundberg?" In: *La Rivista Italiana della Medicina di Laboratorio-Italian Journal of Laboratory Medicine* 13.3-4 (2017), pp. 127–133 (cit. on p. 27).

[10] Lucian Carata et al. "A primer on provenance". In: *Queue* 12.3 (2014), p. 10 (cit. on p. 10).

[11] Paolo Carraro and Mario Plebani. "Errors in a Stat Laboratory: Types and Frequencies 10 Years Later". In: *Clinical Chemistry* 53.7 (2007), pp. 1338–1342. ISSN: 0009-9147. DOI: 10.1373/clinchem.2007.088344. eprint: http://clinchem.aaccjnls.org/content/53/7/1338.full.pdf. URL: http://clinchem.aaccjnls.org/content/53/7/1338 (cit. on p. 12).

[12] CEN. *CEN/TC 140-In vitro diagnostic medical devices*. URL: https://www.din.de/en/getting-involved/standards-committees/named/european-committees/wdc-grem:din21:54739810 (visited on 01/2019) (cit. on p. 11).

[13] Grégoire Chamayour. *History and Philosophy of Traceability*. 2014. URL: https://www.mpiwg-berlin.mpg.de/research/projects/DeptII_ChamayouGregoire-Traceability (visited on 01/2019) (cit. on p. 5).

[14] CLSI Characterization. "qualification of commutable reference materials for Laboratory Medicine: approved guideline". In: *CLSI document C53-A. Wayne, PA: CLSI* (2010) (cit. on p. 10).

[15] International Federation of Clinical Chemistry and Laboratory Medicine. *Advancing excellence in laboratory medicine for better healthcare worldwide*. 2017. URL: http://www.ifcc.org/about/ (visited on 01/2019) (cit. on p. 10).

[16] International Federation of Clinical Chemistry and Laboratory Medicine. *Task Force on the Impact of Laboratory Medicine on Clinical Management and Outcomes (TF-ICO)*. 2017. URL: http://www.ifcc.org/executive-board-and-council/eb-task-forces/task-force-on-the-impact-of-laboratory-medicine-on-clinical-management-and-outcomes-tf-ico/ (visited on 01/2019) (cit. on p. 12).

[17] B Collomb et al. "5PSQ-136 Computerisation of medical devices, traceability and unexpected loss of data: reports and prospects for improvement". In: *European Journal of Hospital Pharmacy: Science and Practice* 25.Suppl 1 (2018). BMJ Specialist Journals, A226–A228.

ISSN: 2047-9956. DOI: 10.1136/ejhpharm-2018-eahpconf.488. eprint: https://ejhp.bmj.com/content/25/Suppl_1/A226.3.full.pdf. URL: https://ejhp.bmj.com/content/25/Suppl_1/A226.3 (cit. on p. 8).

[18] European Commission. *Communication from the commission to the European parliament, the council,the European economic and social committee and the committee of the regionsSafe, effective and innovative medical devices and in vitro diagnostic medi-cal devices for the benefit of patients, consumers and healthcare professionals*. 2012. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52012DC0540 (visited on 01/2019) (cit. on p. 8).

[19] US President's Information Technology Advisory Committee. *Revolutionizing Health Care Through Information Technology*. 2004. URL: https://www.nitrd.gov/pubs/pitac/pitac_report_health-it_2004.pdf (visited on 01/2019) (cit. on p. 9).

[20] Deloitte. *Continuous interconnected supply chain Using Blockchain and Internet-of-Things in supply chain traceability*. 2017. URL: https://www2.deloitte.com/content/dam/Deloitte/lu/Documents/technology/lu-blockchain-internet-things-supply-chain-traceability.pdf (visited on 01/2019) (cit. on p. 9).

[21] DICOM. *DICOM Supplement 122: Specimen Identification and Revised Pathology SOP Classes*. 2006. URL: ftp://medical.nema.org/medical/dicom/final/sup122_ft2.pdf (visited on 01/2019) (cit. on p. 27).

[22] Reinhold Dunkl et al. "Assessing medical treatment compliance based on formal process modeling". In: *Symposium of the Austrian HCI and Usability Engineering Group*. Springer. 2011, pp. 533–546 (cit. on p. 25).

[23] Gregor Engels et al. "Process modeling using UML". In: (Oct. 2005), pp. 83–117 (cit. on p. 16).

[24] Joshua C. Feblowitz et al. "Summarization of clinical information: A conceptual model". In: *Journal of Biomedical Informatics* 44.4 (2011), pp. 688–699. ISSN: 1532-0464. DOI: https://doi.org/10.1016/j.jbi.2011.03.008. URL: http://www.sciencedirect.com/science/article/pii/S1532046411000591 (cit. on p. 10).

[25] US Food and Drugs Administration. *FDA Fundamentals*. 2018. URL: https://www.fda.gov/aboutfda/transparency/basics/ucm192695.htm (visited on 01/2019) (cit. on p. 7).

[26] US Food and Drugs Administration. *FDA MEDICAL DEVICE TRACKING REQUIREMENTS*. 2018. URL: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=821 (visited on 01/2019) (cit. on p. 7).

[27] US Food and Drugs Administration. *Medical Device Tracking Guidance for Industry and Food and Drug Administration Staff*. 2018. URL: https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM071775.pdf (visited on 01/2019) (cit. on p. 8).

[28] US Food and Drugs Administration. *Unique Device Identification - UDI*. 2018. URL: https://www.fda.gov/medicaldevices/deviceregulationandguidance/uniquedeviceidentification/ (visited on 01/2019) (cit. on p. 8).

[29] Forbes. *How Blockchain Is Impacting Healthcare And Life Sciences Today*. 2018. URL: https://www.forbes.com/sites/ciocentral/2018/04/02/how-blockchain-is-impacting-healthcare-and-life-sciences-today/ (visited on 01/2019) (cit. on p. 9).

[30] Forbes. *openEHR Task Planning Model Specification*. 2013-2018. URL: https://www.openehr.org/releases/BASE/latest/docs/architecture_overview/architecture_overview.html (visited on 01/2019) (cit. on pp. 21, 22, 24, 62, 66, 74, 78).

[31] International Medical Device Regulation Forum. *UDI Guidance - Unique Device Identification (UDI) of Medical Devices*. 2013. URL: http://www.imdrf.org/docs/imdrf/final/technical/imdrf-tech-131209-udi-guidance-140901.pdf (visited on 01/2019) (cit. on p. 8).

[32] National Sanitation Foundation. *What systems you need and what to expect from a traceability audit*. 2015. URL: https://www.nsf.org/newsroom_pdf/NewFood_Traceability_Oct2015.pdf (visited on 01/2019) (cit. on p. 5).

[33] openEHR Foundation. *openEHR website*. URL: https://www.openehr.org/ (visited on 01/2019) (cit. on p. 3).

[34] openEHR Foundation. *Who is using openEHR*. 2018. URL: https://www.openehr.org/who_is_using_openehr/ (visited on 01/2019) (cit. on p. 59).

[35] Martin Fowler and Kendall Scott. *UML Distilled (2Nd Ed.): A Brief Guide to the Standard Object Modeling Language*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2000. ISBN: 0-201-65783-X (cit. on p. 16).

[36] Lundberg GD. "Acting on significant laboratory results". In: *JAMA* 245.17 (1981), pp. 1762–1763. DOI: 10.1001/jama.1981.03310420052033. eprint: /data/journals/jama/9115/jama_245_17_033.pdf. URL: +%20http://dx.doi.org/10.1001/jama.1981.03310420052033 (cit. on p. 27).

[37] Orlena CZ Gotel and CW Finkelstein. "An analysis of the requirements traceability problem". In: *Requirements Engineering, 1994., Proceedings of the First International Conference on*. IEEE. 1994, pp. 94–101 (cit. on p. 11).

[38] Object Management Group. *OMG BPMN Specifications - version 2.0 - 2011*. 2011. URL: https://www.omg.org/spec/BPMN/2.0/About-BPMN/ (visited on 01/2019) (cit. on pp. 16, 17).

[39] Object Management Group. *OMG UML Specifications - version 2.5-1 - 2017*. 2017. URL: https://www.omg.org/spec/UML/About-UML/ (visited on 01/2019) (cit. on p. 16).

[40] ISO Guide. *Reference Materials-General and statistical principles for certification. 2006*. 35 (cit. on p. 10).

[41] Joint Committee for Guides in Metrology. *JCGM International vocabulary of metrology*. 2008. URL: https://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2008.pdf (visited on 01/2019) (cit. on p. 9).

[42] Mike Hallworth. "Measuring the impact of laboratory medicine on clinical management and patient outcomes". In: *The Journal of the International Federation of Clinical Chemistry and Laboratory Medicine* 26 (2015), pp. 4–6 (cit. on p. 12).

[43] Mike J Hallworth. "Improving clinical outcomes–towards patient-centred laboratory medicine". In: *Annals of clinical biochemistry* 52.6 (2015), pp. 715–716 (cit. on p. 12).

[44] Mike J. Hallworth et al. "Current Evidence and Future Perspectives on the Effective Practice of Patient-Centered Laboratory Medicine". In: *Clinical Chemistry* 61.4 (2015), pp. 589–599. ISSN: 0009-9147. DOI: 10.1373/clinchem.2014.232629. eprint: http://clinchem.aaccjnls.org/content/61/4/589.full.pdf. URL: http://clinchem.aaccjnls.org/content/61/4/589 (cit. on p. 12).

[45] Douglas Harper. *Online Etymology Dictionary*. 2010. URL: https://www.dictionary.com/browse/traceability (visited on 01/2019) (cit. on p. 5).

[46] Becker's Healthcare. *Blockchain securing the healthcare supply chain.* 2018. URL: https://www.beckershospitalreview.com/supply-chain/blockchain-securing-the-healthcare-supply-chain.html (visited on 01/2019) (cit. on p. 9).

[47] Arthur HM ter Hofstede et al. *Modern Business Process Automation: YAWL and its support environment.* Springer Science & Business Media, 2009 (cit. on p. 16).

[48] Sam Hume et al. "Visualizing and Validating Metadata Traceability within the CDISC Standards". In: *AMIA Summits on Translational Science Proceedings* 2017 (2017), p. 158 (cit. on p. 11).

[49] ICHCLR. *International Consortium for Harmonization of Clinical Laboratory Results.* 2017. URL: https://www.harmonization.net/ (visited on 01/2019) (cit. on p. 10).

[50] GO FAIR Initiative. *R1.2: (Meta)data are associated with detailed provenance.* URL: https://www.go-fair.org/fair-principles/r1-2-metadata-associated-detailed-provenance/ (visited on 01/2019) (cit. on p. 10).

[51] World Resource Institute. *Where do the products come from?* 2016. URL: https://sustainableforestproducts.org/sites/default/files/2016-09/Traceability.pdf (visited on 01/2019) (cit. on p. 5).

[52] CNN Interactive. *Hamburger recall rises to 25 million pounds.* 1997. URL: http://edition.cnn.com/US/9708/21/beef.update/ (visited on 01/2019) (cit. on p. 5).

[53] HL7 International. *FHIR Resource Provenance.* URL: https://www.hl7.org/fhir/provenance.html (visited on 01/2019) (cit. on p. 10).

[54] HL7 International. *HL7 Domain Analysis Model: Specimen, Release 1.* https://www.w3.org/TR/prov-overview/ (last accessed in January 2019) (cit. on p. 35).

[55] IHE International. *IHE Process.* URL: https://www.ihe.net/about_ihe/ihe_process/ (visited on 01/2019) (cit. on p. 18).

[56] IHE International. *Integrating the Healthcare Enterprise website.* URL: https://www.ihe.net (visited on 01/2019) (cit. on pp. 3, 18).

[57] ISO. *ISO 18308:2011 - Health informatics – Requirements for an electronic health record architecture.* 2011. URL: https://www.iso.org/standard/52823.html (visited on 01/2019) (cit. on p. 10).

[58] ISO. *ISO/IEC 15944-9:2015-Information technology — Business operational view — Part 9: Business transaction traceability framework for commitment exchange*. 2015. URL: https://www.iso.org/obp/ui/#iso:std:iso-iec:15944:-9:ed-1:v1:en (visited on 01/2019) (cit. on p. 12).

[59] ISO. *ISO/TC 176/SC 1 22005:2005, Traceability in the feed and food chain e General principles and basic requirements for system design and implementation*. 2005 (cit. on p. 5).

[60] ISO. *ISO/TC 176/SC 1 8402:1994, Quality management and quality assurance e Vocabularys*. 1994 (cit. on p. 5).

[61] ISO. *ISO/TC 176/SC 1 9000:2000, ISO/TC 176/SC 1 9000:2000, Quality management systems e Fundamentals and vocabulary*. 2000 (cit. on p. 5).

[62] ISO. *ISO/TC 276/WG5 Data Processing and integration*. URL: http://globalbiobankweek.org/wp-content/uploads/2017/03/Martin_GOLEBIEWSKI.pdf (visited on 01/2019) (cit. on p. 11).

[63] Graham RD Jones and Craig Jackson. "The Joint Committee for Traceability in Laboratory Medicine (JCTLM)–its history and operation". In: *Clinica Chimica Acta* 453 (2016), pp. 86–94 (cit. on p. 10).

[64] Joris van de Klundert, Pascal Gorissen, and Stef Zeemering. "Measuring clinical pathway adherence". In: *Journal of biomedical informatics* 43.6 (2010), pp. 861–872 (cit. on p. 25).

[65] Common Workflow Language. *Common Workflow Language - version 1.0.2 - 2016*. 2010-2016. URL: https://www.commonwl.org/ (visited on 01/2019) (cit. on p. 17).

[66] Peter Lawrence, ed. *Workflow Handbook 1997*. New York, NY, USA: John Wiley & Sons, Inc., 1997. ISBN: 0-471-96947-8 (cit. on pp. 14, 15).

[67] Lucian L Leape et al. "The nature of adverse events in hospitalized patients: results of the Harvard Medical Practice Study II". In: *New England journal of medicine* 324.6 (1991), pp. 377–384 (cit. on p. 9).

[68] Bela G Liptak and Halit Eren. "Standards in Process Control and Automation". In: *Instrument Engineers' Handbook, Volume 3*. CRC Press, 2016, pp. 210–217 (cit. on p. 10).

[69] Ronny S. Mans et al. "Process Mining in Healthcare: Data Challenges when Answering Frequently Posed Questions". In: *Proceedings of the 2012 International Conference on Process Support and Knowledge Representation in Health Care*. BPM' 2012. Tallinn, Estonia: Springer-Verlag, 2013, pp. 140–153. ISBN: 978-3-642-36437-2. DOI: 10.1007/978-3-642-36438-9_10. URL: http://dx.doi.org/10.1007/978-3-642-36438-9_10 (cit. on p. 25).

[70] Ronny S Mans, Wil MP Van der Aalst, and Rob JB Vanwersch. *Process mining in healthcare: Evaluating and exploiting operational healthcare processes*. Springer, 2015 (cit. on p. 25).

[71] Maria Martinez Perez, Carlos Dafonte, and Angel Gomez. "Traceability in Patient Healthcare through the Integration of RFID Technology in an ICU in a Hospital". In: *Sensors* 18.5 (2018), p. 1627 (cit. on p. 12).

[72] Maria Martinez Perez, Guillermo Vazquez Gonzalez, and Carlos Dafonte. "Safety and traceability in patient healthcare through the integration of RFID technology for intravenous mixtures in the prescription-validation-elaboration-dispensation-administration circuit to day hospital patients". In: *Sensors* 16.8 (2016), p. 1188 (cit. on p. 9).

[73] Laura Maruster and René J Jorna. "From data to knowledge: a method for modeling hospital logistic processes". In: *IEEE Transactions on information technology in biomedicine* 9.2 (2005), pp. 248–255 (cit. on p. 25).

[74] Luc Moreau et al. "The Open Provenance Model core specification (v1.1)". In: *Future Generation Computer Systems* 27.6 (2011), pp. 743–756. ISSN: 0167-739X. DOI: https://doi.org/10.1016/j.future.2010.07.005. URL: http://www.sciencedirect.com/science/article/pii/S0167739X10001275 (cit. on p. 10).

[75] Luc Moreau et al. "The rationale of PROV". In: *Web Semantics: Science, Services and Agents on the World Wide Web* 35 (2015), pp. 235–257. ISSN: 1570-8268. DOI: https://doi.org/10.1016/j.websem.2015.04.001. URL: http://www.sciencedirect.com/science/article/pii/S1570826815000177 (cit. on p. 10).

[76] Petter Olsen and Melania Borit. "How to define traceability". In: *Trends in Food Science and Technology* 29.2 (2013), pp. 142–150. ISSN: 0924-2244. DOI: https://doi.org/10.1016/j.tifs.2012.10.003. URL: http://www.sciencedirect.com/science/article/pii/S0924224412002117 (cit. on p. 5).

[77] openEHR. *openEHR Architecture Overview*. 2013-2018. URL: https://www.openehr.org/releases/BASE/latest/docs/architecture_overview/architecture_overview.html (visited on 01/2019) (cit. on p. 61).

[78] OPM. *Open Provenance*. URL: https://openprovenance.org/ (visited on 01/2019) (cit. on p. 10).

[79] World Health Organization. *Human organ and tissue transplantation*. 2010. URL: http://apps.who.int/gb/ebwha/pdf_files/WHA63/A63_R22-en.pdf (visited on 01/2019) (cit. on p. 7).

[80] EU Parliament. *Directive 98/79/EC of the European Parliaments and of the Council of 27 October 1998 on in vitro diagnostic medical devices*. 1998. URL: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52007PC0355 (visited on 01/2019) (cit. on p. 10).

[81] European Parliament. *Regulation (EC) No 178/2002 of the European Parliament and of the Council of 28 January 2002 laying down the general principles and requirements of food law, establishing the European Food Safety Authority and laying down procedures in matters of food safety. Official Journal of the European Communities, 31, 1e24*. 2002. URL: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A32002R0178 (visited on 01/2019) (cit. on p. 5).

[82] IHE Pathology and Laboratory Medicine Committee. *IHE PaLM Technical Frameworks*. URL: https://www.ihe.net/resources/technical_frameworks/#PaLM (visited on 01/2019) (cit. on pp. 18, 20).

[83] Maria Martinez Perez, Guillermo Vazquez Gonzalez, and Carlos Dafonte. "Safety and Traceability in Patient Healthcare through the Integration of RFID Technology for Intravenous Mixtures in the Prescription-Validation-Elaboration-Dispensation-Administration Circuit to Day Hospital Patients". In: *Sensors* 16.8 (2016). ISSN: 1424-8220. DOI: 10.3390/s16081188. URL: http://www.mdpi.com/1424-8220/16/8/1188 (cit. on p. 7).

[84] Maria Martinez Perez, Guillermo Vazquez Gonzalez, and Carlos Dafonte. "The development of an RFID solution to facilitate the traceability of patient and pharmaceutical data". In: *Sensors* 17.10 (2017), p. 2247 (cit. on p. 7).

[85] Mario Plebani. "Diagnostic errors and laboratory medicine–causes and strategies". In: *EJIFCC* 26.1 (2015), p. 7 (cit. on p. 12).

[86] Mario Plebani. "The detection and prevention of errors in laboratory medicine". In: *Annals of clinical biochemistry* 47.2 (2010), pp. 101–110 (cit. on p. 12).

[87] Mario Plebani, Michael Laposata, and George D. Lundberg. "The Brain-to-Brain Loop Concept for Laboratory Testing 40 Years After Its Introduction". In: *American Journal of Clinical Pathology* 136.6 (2011), pp. 829–833. DOI: 10.1309/AJCPR28HWHSSDNON. eprint: /oup/backfile/content_public/journal/ajcp/136/6/10.1309_ajcpr28hwhssdnon/3/ajcpath136-0829.pdf. URL: http://dx.doi.org/10.1309/AJCPR28HWHSSDNON (cit. on pp. 27, 28).

[88] Oxford University Press. *Oxford Living Dictionaries*. URL: https://en.oxforddictionaries.com/definition/traceable (visited on 01/2019) (cit. on p. 5).

[89] Gilbert Regan et al. "Medical device standards' requirements for traceability during the software development lifecycle and implementation of a traceability assessment model". In: *Computer Standards & Interfaces* 36.1 (2013), pp. 3–9 (cit. on p. 11).

[90] Victor M Alonso Roris et al. "Towards a mobile-based platform for traceability control and hazard analysis in the context of parenteral nutrition: description of a framework and a prototype app". In: *JMIR research protocols* 5.2 (2016) (cit. on p. 12).

[91] Matthew Rubinstein et al. "Effectiveness of Practices to Support Appropriate Laboratory Test Utilization: A Laboratory Medicine Best Practices Systematic Review and Meta-Analysis". In: *American journal of clinical pathology* 149.3 (2018), pp. 197–221 (cit. on p. 12).

[92] GS1 standards. *Healthcare traceability and GS1 standards*. 2013. URL: https://www.gs1.org/traceability-healthcare (visited on 01/2019) (cit. on p. 8).

[93] A. Sulis et al. "Process Mining Applied to Laboratory Workflow: A Comparison between Two Sites". In: *2015 International Conference on Healthcare Informatics (ICHI)*. Vol. 00. Oct. 2015, pp. 437–443. DOI: 10.1109/ICHI.2015.61. URL: doi.ieeecomputersociety.org/10.1109/ICHI.2015.61 (cit. on p. 13).

[94] A. Sulis et al. "Traceability Based Description of Clinical Processes: Extension of IHE Guidelines for Phlebotomy Workflows". In: *EJBI - European Journal for Biomedical Informatics* 12 (2016), en36–en44. URL: http://publications.crs4.it/pubdocs/2016/SFMGDTGCZ16 (cit. on p. 58).

[95] Joint Committee for Traceability in Laboratory Medicine. *Traceability in Laboratory Medicine: Important for Patients*. 1998. URL: https://www.jctlm.org/ (visited on 01/2019) (cit. on p. 10).

[96] P. Tracol. "Materials vigilance and traceability". In: *Orthopaedics and Traumatology: Surgery and Research* 102.1, Supplement (2016). 2015 Instructional Course Lectures (SoFCOT), S95–S103. ISSN: 1877-0568. DOI: https://doi.org/10.1016/j.otsr.2015.05.013. URL: http://www.sciencedirect.com/science/article/pii/S1877056815003527 (cit. on pp. 8, 9).

[97] Wil Van Der Aalst et al. "Process mining manifesto". In: *International Conference on Business Process Management*. Springer. 2011, pp. 169–194 (cit. on p. 23).

[98] Hubert W Vesper, Gary L Myers, and W Greg Miller. "Current practices and challenges in the standardization and harmonization of clinical laboratory tests–". In: *The American journal of clinical nutrition* 104.suppl_3 (2016), 907S–912S (cit. on p. 10).

[99] Hubert W Vesper and Linda M Thienpont. "Traceability in laboratory medicine". In: *Clinical chemistry* 55.6 (2009), pp. 1067–1075 (cit. on p. 10).

[100] W3C. *An Overview of the PROV Family of Documents*. URL: https://www.w3.org/TR/prov-overview/ (visited on 01/2019) (cit. on pp. 10, 11).

[101] Corinne Wales, Mark Harvey, and Alan Warde. "Recuperating from BSE: The shifting UK institutional basis for trust in food". In: *Appetite* 47.2 (2006), pp. 187–195. ISSN: 0195-6663. DOI: https://doi.org/10.1016/j.appet.2006.05.007. URL: http://www.sciencedirect.com/science/article/pii/S0195666306003886 (cit. on p. 5).

[102] WPI. *Workflow Pattern Initiative*. 2010-2017. URL: http://www.workflowpatterns.com/ (visited on 01/2019) (cit. on p. 17).

APPENDIX

Specimen Exent Tracking Profile
Detailed Events-Metadata matrix

Event	Attribute	Data type	Description	Card	Usage	DAM Mapping
Specimen Collected	Placer Order Number	String	PON as usual in HL7 v2	0…*	RE	No mapping, use HL7 v2 instead?
	Collector	String /ID	Person responsible of Specimen Collection	1…1	R	Performer.identifier
	Type	String/Code	Specimen Type as usual in HL7 V2	0…1	O	Specimen.typeCode
	Form	String/Code	Material nature (i.e., liquid, gas)	0…1	O	Specimen.formCode
	Description	String	Additional specimen information	0…1	O	Specimen.description
	Procedure	String	Activity performed for collection (i.e., venipuncture, biopsy)	0…1	O	SpecimenCollectionProcedure.ProcedureCode
	Coll. Date Range	Timestamp (range)	Time range of collection duration	0…1	C	SpecimenCollectionProcedure.actualCollectionDateRange
	Missed Reason?	String	Reason of collection not completed	1…1	C	SpecimenCollectionProcedure.missedReason
	Container Name	String	Name (model?) of the container	0…1	RE	SpecimenContainer.name
	Container Material	String/Code	Material of the specimen container	0…1	O	SpecimenContainer.containerMaterialCode
	Container Capacity	Number/Code	Capacity of the specimen Container	0…1	O	SpecimenContainer.Parameters.capacity
	Container Additive	String/Code	Additive of the specimen Container	0…1	O	SpecimenContainer.additive
	Container Rank	Number	Rank of the container collecting the specimen (1=first, 2= second, and so on)	0…1	RE	
	Container Height	Number	Height of the specimen container	0…1	O	SpecimenContainerParameters.height
	Expiration Time	Timestamp	Date after the specimen is no longer viable	1…1	R	Specimen.expirationTime
Specimen Containers Prepared	Placer Order Number	String	PON as usual in HL7 v2	0…*	RE	No mapping, use HL7 v2 instead?
	Producer	String /ID	Producer responsible of Specimen labeled containers production	0…1	RE	Performer.identifier,
	Producer	String	Prodycer type (Human, Robotic System)	0…1	RE	Performer.typeCode
	Container Name	String	Name (model?) of the container	0…1	RE	SpecimenContainer.name
	Container Material	String/Code	Material of the specimen container	0…1	O	SpecimenContainer.containerMaterialCode
	Container Capacity	Number/Code	Capacity of the specimen Container	0…1	O	SpecimenContainer.Parameters.capacity
	Container Additive	String/Code	Additive of the specimen Container	0…1	O	SpecimenContainer.additive
	Container Rank	Number	Overall number of containers where the specimen has been collected	0…1	RE	
	Container Height	Number	Height of the specimen container	0…1	O	SpecimenContainerParameters.height
	Accept Entity	String/ID	Location where the specimen has been accepted	1…1	R	SpecimenMoveActivity.toEntity

APPENDIX

Specimen Exent Tracking Profile
Detailed Events-Metadata matrix

Event	Attribute	Data type	Description	Card	Usage	DAM Mapping
Specimen Accepted	Accept Timestamp	Timestamp	Timestamp of acceptance	1…1	R	
Specimen Rejected	Reject Entity	String/ID	Location where the specimen has been rejected	1…1	R	SpecimenMoveActivity.toEntity
	Reject Timestamp	Timestamp	Timestamp of rejection of the specimen	1…1	R	
	Reject Reason	String	Reason why the specimen has been rejected	1…1	R	SpecimenMoveActivity.varianceReasonCode
Specimen Re-identified	New Specimen identifier	String/ID	New specimen identifier	1…1	R	Specimen.specimenIdentifier
	New Container Identifier	String/ID	New container identifier	1…1	R	SpecimenContainer.containerIdentifier
	Re-identification Entity	String/ID	Location where the specimen has been re-identified	1…1	R	SpecimenMoveActivity.toEntity
	Is De-identified	Boolean	Boolean that explains a de-identification	1…1	R	
Specimen Archived	Expiration Time	Timestamp	Date after the specimen is no longer viable	1…1	R	Specimen.expirationTime
	Original Measurement	Number	Initial volume (i.e., container capacity) of the specimen	1…1	R	Specimen.originalSpecimenMeasurement
	Current Measurement	Number	Current volume of the specimen	1…1	R	Specimen.currentSpecimenMeasurement
	Current Status	String	Status of the specimen at the time of archiving	1…1	R	SpecimenCollectionProcedure.statusCode
Specimen Retrieved	Expiration Time	Timestamp	Date after the specimen is no longer viable	1…1	R	Specimen.expirationTime
	Original Measurement	Number	Initial volume (i.e., container capacity) of the specimen	1…1	R	Specimen.originalSpecimenMeasurement
	Current Measurement	Number	Current volume of the specimen	1…1	R	Specimen.currentSpecimenMeasurement
	Retriever Name	String	Name of the person responsible for retrieve	1…1	R	
	Retriever Identifier	String/ID	Identifier of the person responsible for retrieve	1…1	R	
	Retriever Location	String	Location where the retrieve has been performed	1…1	R	
	Reason Retrieve	String	Reason for specimen retrieve	1…1	R	
	Involved Specialty	String/Code	Specialty related to specimen retrieve	0…1	O	No mapping, use HL7 v2 instead?
	Involved Diagnosis	String/Code	Diagnosis related to specimen retrieve	0…1	O	No mapping, use HL7 v2 instead?
Specimen Derived	Parent Identifier	String/ID	Parent ID of the specimen from where the specimen has been derived	1…1	R	Specimen.specimenIdentifier
	Type	String	Specimen Type as usual in HL7 V2	1…1	R	Specimen.typeCode
	Current Measurement	Number	Volume of the derived specimen	0…1	O	Specimen.currentSpecimenMeasurement

APPENDIX

Specimen Exent Tracking Profile
Detailed Events-Metadata matrix

Event	Attribute	Data type	Description	Card	Usage	DAM Mapping
	Original Measurement	Number		0…1	O	Specimen.originalSpecimenMeasurement
	Specimen Child Role	String	Role of the derived specimen (i.e., aliquot, block for tissue)	0…1	O	Specimen.childRole
Specimen Processing Start	Processing Procedure	String	Description of the processing step started	0…1	O	SpecimenProcessingActivity.processingProcedure
	Processing Additive	String	Substance required and added to the specimen for processing	0…1	O	SpecimenProcessingActivity.processingAdditive
	Temperature	Number	Temperature at which the processing occurred	0…1	O	SpecimenProcessingActivity.Temperature
Specimen Processing End	Processing Procedure	String	Description of the processing step ended	0…1	O	SpecimenProcessingActivity.processingProcedure
	Processing Additive	String	Substance required and added to the specimen for processing	0…1	O	SpecimenProcessingActivity.processingAdditive
	Temperature	Number	Temperature at which the processing occurred	0…1	O	SpecimenProcessingActivity.Temperature
Specimen Arrived at Location	To Entity	String/ID	Location where the specimen arrived	1…1	R	SpecimenMoveActivity.toEntity
	Parent Package ID	String/ID	Main parent package container	0…1	O	
	Package ID	String/ID	Package container	0…1	O	
	Position in Parent Package	String/ID	Position assumed in parent package	0…1	O	
Specimen Left Location	From Entity	String/ID	Location from where the specimen has been transferred	1…1	R	SpecimenMoveActivity.FromEntity
	Parent Package ID	String/ID	Main parent package container	0…1	O	
	Package ID	String/ID	Package container	0…1	O	
	Position in Parent Package	String/ID	Position assumed in parent package	0…1	O	
Specimen Discarded	Discarding Timestamp	Timestamp	Timestamp of specimen discarding	1…1	R	
	Discarding Reason	String	Reason for specimen discarding	1…1	R	